

VU Research Portal

Does exposure to facial composites damage eyewitness memory?

Sporer, Siegfried L.; Tredoux, Colin G.; Vredeveldt, Annelies; Kempen, Kate; Nortje, Alicia

published in

Applied Cognitive Psychology
2020

DOI (link to publisher)

[10.1002/acp.3705](https://doi.org/10.1002/acp.3705)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Sporer, S. L., Tredoux, C. G., Vredeveldt, A., Kempen, K., & Nortje, A. (2020). Does exposure to facial composites damage eyewitness memory? A comprehensive review. *Applied Cognitive Psychology*, 34(5), 1166-1179. <https://doi.org/10.1002/acp.3705>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



RESEARCH ARTICLE

WILEY

Does exposure to facial composites damage eyewitness memory? A comprehensive review

Siegfried L. Sporer¹ | Colin G. Tredoux^{2,3} | Annelies Vredeveldt⁴ |
Kate Kempen² | Alicia Nortje²

¹Department of Psychology and Sports Science, University of Giessen, Giessen, Germany

²Department of Psychology, University of Cape Town, Cape Town, South Africa

³CLLE, Université de Toulouse, CNRS, UT2J, Toulouse, France

⁴Department of Criminal Law and Criminology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence

Siegfried L. Sporer, Department of Psychology, University of Giessen, Giessen, Germany.
Email: siegfried.l.sporer@psy.jlug.de

Funding information

Deutsche Forschungsgemeinschaft (Sp272/9-1)

Summary

Eyewitnesses often create face likenesses, which are published in the hope that potential suspects will be reported to the police. Witnesses exposed to another witness's composite, however, may be positively or negatively influenced by such composites. A good likeness may facilitate identification, but a bad likeness that resembles an innocent suspect may lead to a misidentification (“mix-up”). We offer a theoretical review, and comprehensively summarize extant studies descriptively because most studies did not report enough statistical details to warrant a formal meta-analysis. Some studies showed negative exposure effects, particularly when the innocent suspect and composite shared misleading features. Studies that exposed witnesses to “good” composites reported positive or no effects on lineup performance, and some highly powered studies also showed no effect. We outline suggestions for further investigations under ecologically valid conditions. We also make recommendations for investigative practice, and the evaluation of identification evidence by fact finders or courts.

KEYWORDS

eyewitness identification, eyewitness recall, face composites, misinformation effect

1 | INTRODUCTION

Mistaken eyewitness identifications can have a major impact on criminal cases, and there is plenty of evidence that mistaken identifications have been involved in wrongful convictions discovered through DNA testing (Garrett, 2011; www.innocenceproject.org). One potential factor contributing to these misidentifications may be the production of face composites as part of the police investigation. In cases in which the identity of the perpetrator is at issue, the police may rely on eyewitnesses to help produce a likeness (or “composite”) of the perpetrator's facial appearance. By publishing the composite, police investigators hope that a member of the public will recognize the person depicted and report this to the police (Davies & Valentine, 2007; Shepherd & Ellis, 1996). Face composites have been used by the police across the globe for many decades, from the US to Europe and Australia and many other countries, including Brazil, South Africa and

in former socialist and communist countries (Saraiva et al., 2018; Schmidt & Tredoux, 2006; Shepherd & Ellis, 1996; Snetkow, Sinin, & Delang, 1981).

Although face composites are commonly used by police, laboratory research indicates that composites often do not resemble the perpetrator (Davies & Valentine, 2007; Frowd et al., 2005; Frowd et al., 2005; Koehn & Fisher, 1997; Kovera, Penrod, Pappas, & Thill, 1997). However, Frowd, Valentine, and Davis (2015) summarize data on new generation composite systems¹ and provide evidence that resemblance between composites and targets may be improving. Newer interviewing techniques developed specifically for face composite construction may further improve composite accuracy (Skelton et al., 2019; Fodarella et al., 2017; Fodarella, Kuivaniemi-Smith, Gawrylowicz, & Frowd, 2015).

A facial composite constructed by a witness can affect later identification in at least two ways. First, the task of constructing a

composite may affect that witness's later ability to identify the true perpetrator from a lineup (*composite construction effect*). Second, exposure to a composite constructed by one witness may affect another witness's subsequent lineup identification decision who followed the case in the media (*composite exposure effect*). When we set out to review the literature on construction and exposure effects on later identification, we had planned to conduct separate meta-analyses on each of these effects. Unfortunately, after identifying relevant studies, it became clear that conducting a meta-analysis for the composite exposure effect would be indefensible. This was due to the fact that (a) there was a small number of studies to begin with, (b) several studies did not report sufficient information (e.g., cell means and cell sizes) to calculate effect sizes and the authors informed us that those data were no longer available, and (c) several studies collapsed the data across different types of control conditions (involving no composite and a good-quality composite, respectively). We therefore report our meta-analysis of the composite construction effect in a companion article (Tredoux, Sporer, Vredeveltdt, Kempen, & Nortje, 2020) and present a narrative comprehensive review of the composite exposure effect in the present article.

2 | POTENTIAL CONSEQUENCES OF COMPOSITE CONSTRUCTION AND EXPOSURE

The construction of a face composite may affect a later identification attempt in a lineup in at least two ways. Figure 1 gives a schematic overview of a possible course of events in a hypothetical case with two witnesses, A and B. Both witnesses observe a crime and provide a description of the perpetrator after a short delay (Delay 1a—this kind of delay is usually quite short, since the first report to the police is often on the same day, or in the context of a more thorough interview on the following day [see the archival analysis by Sporer, 1992a]).

Witness A, after an additional delay after the person description (Delay 1b) will meet with a police artist or a composite construction expert, who (again) asks the witness for a description, based on which a composite is constructed. This composite is then distributed among

the police and/or published so that members of the public can view it. If the composite is of high quality, that is, if it bears a strong resemblance to the true perpetrator, someone may notice the likeness and report the person depicted in the composite, which in turn may lead to an arrest of the true perpetrator. If the composite has some resemblance to an innocent suspect, that person may be pursued instead. It is also possible that the police will search their files for potential suspects matching the description and composite of Witness A. In any of these cases, the police may then put the suspect in a live or photo lineup for identification. Objectively, this lineup may contain the perpetrator (target-present [TP]) or an innocent suspect (target-absent [TA]). Only in experimental studies, however, can the accuracy of an identification decision be objectively established.

In the same hypothetical case, a second Witness (B) may also provide a description but not attempt to construct a face composite. Witness B may follow the case in the media, and consequently may be exposed to the composite created by Witness A. If Witness B subsequently participates in a lineup, this exposure may affect Witness B's lineup decision (*composite exposure effect*). In this article we provide a comprehensive *narrative* review of nine published and one unpublished study on such potential composite exposure effects. Although we had planned to conduct a formal meta-analysis, too many of these studies did not report sufficient data (e.g., exact percentages, *ns* per cell), or only reported graphic representations of results without statistical details, thus not allowing us to calculate effect sizes, essentially *contra-indicating* a meta-analysis.

A second way in which a composite may affect a later identification attempt is when the same Witness (A) first describes and helps in constructing a composite and subsequently participates in a lineup task (*composite construction effect*; see Figure 1). We will not address this issue here, since it has been reviewed in a recent meta-analysis by Tredoux et al., 2020.² Of course, more complex case constellations are conceivable, for example, when both Witnesses A and B (and/or additional witnesses) construct composites and are, or are not, exposed to each other's composites. Some researchers have also addressed the interesting possibility whether or not an integration of several composites (e.g., by morphing) leads to a higher resemblance of the compound composite than its constituents to the original target image (see Bruce, Ness, Hancock, Newman, & Rarity, 2002; Hasel &

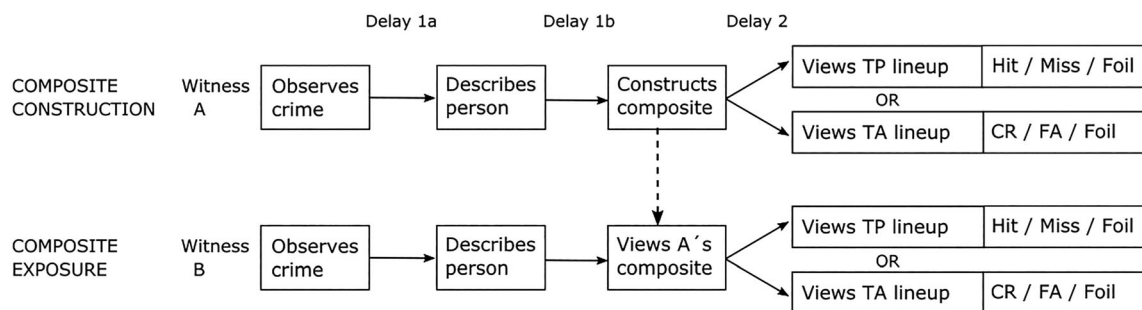


FIGURE 1 Effects of Constructing and/or Exposure to a Composite on Performance in a Person Identification Task. Hit = correct identification of target. FA = incorrect identification of innocent suspect. CR = correct rejection of lineup. Foil = identification of known innocent lineup foil/filler. Miss = incorrect rejection of lineup

Wells, 2007). Here we address only the second scenario, that is the case of Witness B being exposed to a composite created by Witness A.

3 | THEORETICAL APPROACHES TO COMPOSITE EXPOSURE EFFECTS

As far as we know, practically all research on composite exposure effects starts with the assumption that a composite (by Witness A) contains misleading information that will *negatively* affect another Witness's (B) memory, and consequently impair B's description or identification of the perpetrator. However, one should also take the opposite effect into consideration: that exposure to A's composite may *improve* B's performance on a lineup task, perhaps depending on the quality of the composite and the timing of the exposure along the timeline (see Figure 1). In the following manuscript, we will consider both possibilities. When A's composite is viewed shortly after the original observation at the crime scene, discrepancies with Witness B's own memory may be discovered, and hence the composite may be disregarded by the witness (to the extent that this is consciously possible). If exposure to a composite occurs much later, and closer in time to the identification task, the image of A's composite may be more accessible to conscious retrieval (and thus more influential) than the image of the perpetrator seen by B during the crime. Hence, Witness B may pick the person in the lineup who most closely resembles the composite.

3.1 | Misinformation effects from composite viewing

Many studies on misinformation effects distinguish between verbal and visual misinformation. Here, we also distinguish between studies that show effects on *verbal memory*, that is on person descriptions and recall of the course of events, and *visual memory*, by which we mean effects on recognition tasks of objects or person identification lineups. Depending on the form of testing (free recall, open-ended or closed questions, alternative-forced-choice or multiple-choice recognition), and on the modality of testing (verbal or visual), different outcomes have been obtained. These will be outlined below.

Since the 1970s, there have been hundreds of studies that have demonstrated that misleading verbal information or suggestive questioning may lead to a deterioration of memory for event details as well as for memory of the appearance of the person, including misidentifications in a lineup (see Loftus & Greene, 1980; for a review of 30 years of research, see Loftus, 2005). Most studies on composite exposure effects were conducted within the tradition of work on misleading post-event information and its effects on memory, framing the viewing of composites as an example of misinformation that creates a competing memory, or one that updates, or even "overwrites" the original memory (Jenkins & Davies, 1985; Sporer, 1996; Wells, Charman, & Olson, 2005). Less consideration has been given to the

possibility that a "good" composite, that is one with a high resemblance to the target/perpetrator, may strengthen the memory trace, and thus improve later recognition.

Although most misinformation research has investigated the factors that make verbal post-event suggestion a powerful source of misinformation (see Davis & Loftus, 2007), there are at least two paradigms of related research on *visual sources of misinformation*: studies of mugshot exposure effects on later identification, and studies of unconscious transference (Deffenbacher, Bornstein, & Penrod, 2006). In both cases, witnesses mistake a visual image of a face they have seen somewhere else for the face of the perpetrator in a later identification task.

Note that misinformation and mugshot studies were set up explicitly to test whether mistakes in memory reporting can be induced by planting misinformation, whereas the focus in composite exposure studies is whether memory reporting will be affected as a matter of course by viewing a composite image. Here we need to distinguish whether the composite accurately reflects the perpetrator's appearance or contains erroneous features. If it is a good likeness, that is, a good quality composite, the composite may help a second witness make a better identification. If it is a bad likeness of the perpetrator but resembles an innocent individual, who is therefore included by the police as a suspect in a lineup, the second witness might mistakenly identify that person as the perpetrator.

A popular explanation for the interference of face composites and faces originally encoded is "source monitoring failure" (Johnson, Hashtroudi, & Lindsay, 1993). The witness encodes both the original face and the intervening stimulus (e.g., mugshot or composite), and at test has the problem of attributing two competing memories to the appropriate source. If the source of the memories is not clear, witnesses may decide based on a feeling of familiarity, and the intervening stimulus may well seem more familiar because it occurred later in the timeline. Also, the composite may be viewed repeatedly, hence its influence may be stronger than the initial encoding of the perpetrator. Because our memory system is not good at keeping track of why something is familiar (Gronlund & Carlson, 2014), misplaced familiarity may contribute similarly to mugshot exposure effects as well as to composite exposure effects.

3.2 | Change in response criterion

Being exposed to a composite constructed by another witness may not only affect memory accuracy but also a witness's response criterion. Hypothetically, if a witness views a composite that appears to be a good match to the perpetrator, thus confirming her or his memory, the witness may be more likely to choose someone from the lineup. On the other hand, in line with Clare and Lewandowsky's (2004) explanation of the verbal overshadowing effect, exposure to a bad match composite³ may make a witness doubt their own memory and less likely to choose someone. For instance, in Experiment 1 in Wells et al. (2005), 50% of participants in the composite viewing condition made no selection in target-present lineups, compared to only 10% in the control condition, which can be interpreted as the adoption of a more conservative

response criterion. Unfortunately, this study did not contain a target-absent condition, so we do not know if witnesses also adopt a more cautious decision criterion when the perpetrator is not in the lineup.

4 | THE TYPICAL RESEARCH PARADIGM

Experiments that examine the effects of face composites on eyewitness memory are typically conducted in three stages. In the first stage the participant is shown a face—for example a still image of a face, a moving image from a video, or a live person observed carrying out some action (usually a simulated crime like a theft). After a variable period (Delay 1a plus Delay 1b in Figure 1 above), participants in the experimental group construct a face composite or view a face composite that has been constructed by another person. The control group works on a filler task or provides a verbal description of the face. After a further period (Delay 2), a recognition task is administered, in which participants attempt to identify the target from a photo or video lineup or a showup. Tables 1 and 2 summarize these and other design specifics in the studies reviewed, such as the type of composite program used.

5 | REVIEW OF INDIVIDUAL STUDIES

We conducted an extensive search of the literature to find studies on composite exposure and composite construction effects for the present narrative review and the companion meta-analytic review, respectively. We entered the following search terms in various scholarly databases (PsycINFO, MEDLINE, the Social Science Citation Index, Google Scholar): “face composite,” “facial composite,” “face reconstruction,” and “face likeness” as well as names of individual systems like “identikit,” “identi-kit,” “photofit,” “FACES,” “E-FIT,” “Evo-FIT,” and so forth. The database search was supplemented by consulting reference lists of all relevant articles, as well as contacting researchers in the field to request

unpublished manuscripts, conference proceedings, student dissertations, and other types of publications on this topic. Finally, the Thomson Reuters and Google Scholar databases were used to find relevant studies. Table 1 provides a summary of study and design characteristics of the studies reviewed. Notably, most studies used older composite construction systems or programs, and we could not find any newer studies using holistic composite construction systems.

We present a summary of the outcomes of these studies in Tables 3 and 4. Results are separated for TP lineups (nine experiments) and TA lineups (three experiments). Because several studies did not report data separately for hits, identifications of fillers (i.e., known innocent foils), false rejections, or false identifications of the designated misleading filler (here referred to as a “mix-up”), we present the data as best we can, with specifics of individual studies noted in our narrative summary below.

We located 10 experiments (nine published, one unpublished) that investigated composite exposure effects.⁴ Five experiments used a verbal recall test (recall of facial features on a checklist⁵), and 10 used lineups as recognition tests (eight published, one unpublished). Because most of the identification studies reported outcomes differently or in insufficient detail, we had to reconstruct some relevant means (e.g., by averaging across subcategories of the design). Consequently, we could not conduct a formal meta-analysis, as explained earlier.

In the following section, we review individual studies of composite exposure effects (see the overview in Tables 1 and 2).

5.1 | Effects on face recall

In two experiments, Jenkins and Davies (1985) presented participants with a videotape of a classroom intruder (Exp. 1) or of one of two shop lifting incidents (Exp. 2). Composites were created by experienced composite operators and their accuracy was evaluated in pilot studies. The authors showed that viewing composites created by another witness that included incorrect (“misleading”) details not

TABLE 1 Design features and variables manipulated in the studies using recall tasks

Authors (year)	N	Composite system/ program	Comparison groups	Target stimuli	Design specifics	Recall task
Jenkins & Davies (1985, Exp. 1)	97	Photofit; experienced operators	No composite vs. accurate composite vs. altered hair vs. added moustache composite	Film	Two types of misinformation	Recall of facial features on checklist
Jenkins & Davies (1985, Exp. 2)	210	Photofit; experienced operators	No composite vs. accurate composite vs. altered hair vs. added moustache composite	Film	Two types of misinformation	Recall of facial features on checklist
Davies & Jenkins (1985, Exp. 1) ^a	X	Photofit; experienced operators	CG/accurate vs. misleading	Film	Immediate vs. delayed recall	Recall of facial features on checklist
Davies & Jenkins (1985, Exp. 2) ^a	X	Photofit; experienced operators	CG/accurate vs. misleading	Film	Prior recall vs. no recall	Recall of facial features on checklist
Gibling & Davies (1988)	194	Photofit; experienced operators	No composite vs. accurate composite vs. altered hair vs. added moustache composite	Film	Guided memory interview	Recall of facial features on checklist

Note: X = not reported. There were no exact cell means, ns, or inferential statistics reported for these studies.

TABLE 2 Design features and variables manipulated in the studies using lineup identification tasks

Authors (year)	N	Composite system/ program	Comparison groups	Target stimuli	Design specifics	Target presence	Lineup size	Mis- leading foils	Lineup fairness		
Jenkins & Davies (1985, Exp. 2)	210	Photofit; experienced operators	CG/accurate vs. misleading	Film		TP only	12	3	X		
Davies & Jenkins (1985, Exp. 1) ^a	X	Photofit; experienced operators	CG/accurate vs. misleading	Film	Immediate vs. delayed recall	TP only	12	3	X		
Davies & Jenkins (1985, Exp. 2) ^a	X	Photofit; experienced operators	CG/accurate vs. misleading	Film	Prior recall vs. no recall	TP only	12	3	X		
Franzen & Sporer (1994a)	154	Mac-a-mug pro; no experienced operators	CG/accurate vs. misleading	Film	Highly homogeneous lineup	TP and TA	6	1	Several pilot studies		
Franzen & Sporer (1994b)	165	Mac-a-mug pro; no experienced operators	CG/accurate vs. misleading	Film	More heterogeneous lineup	TP and TA	6	1	Several pilot studies		
Dekle (2006)											
Lineups excluding do not know	125	X			No composite vs. unbiased vs. biased composite	Slides		TP and TA	6	0/1 ^a	Pilot study
Showups excluding do not know	98	X			No composite vs. unbiased vs. biased composite	Slides	Separate TA Showups and biased TA Showups	TP and TA	1	0/1 ^a	Pilot study
Wells et al. (2005, Exp. 1)	100	FACES 3.0; no experienced operators; no practice session	No composite vs. composite	Photo	Large number of targets; poor likeness	TP only	6	0	Matching general description		
Kempen (2009, unpublished thesis)	81	FACES 4.0; no experienced operators; practice session of 20 min with participants	No composite vs. composite	Photo		TP only ^b	6	0	Pilot study		

Note: X = not reported. There were no exact cell means, ns, or inferential statistics reported for these studies.

Abbreviations: TA, Target absence; TP, Target presence.

^aIn the unbiased condition, the composite resembled all foils equally; in the biased condition, the composite resembled a particular foil, not the perpetrator (see text).

^bThere was also a TA condition which was presented after the TP condition. Therefore, it will not be used in our analyses.

originally observed in the perpetrator's face, like a moustache or curly (instead of straight) hair, led another witness to also incorporate these details in their recall of the perpetrator's face (Jenkins & Davies, 1985, Exp. 1 and 2). Control participants or those exposed to accurate composites did not make those errors. In both experiments, simply being exposed to a rather good composite did not affect recall performance compared to the no-composite control groups.

In Exp. 1 the whole procedure from target exposure, composite exposure, and memory tests occurred in less than an hour, but in Exp. 2 an immediate (after 20 min) composite exposure and testing group was compared to two delay conditions of 2 or 7 days after which the composites were introduced and the participants' memory tested. However, delay did not affect the results in any way.

In two follow-up experiments described only briefly in conference proceedings, Davies and Jenkins (1985) also included both recall measures and lineup tasks, using the same targets but using a new, longer video of the shoplifting scenario. The procedures and analyses were identical. Thus, a direct comparison with the results of the Jenkins and Davies (1985) study was possible and reported. As expected, in Exp. 1 the prolonged exposure to the target increased accuracy on the facial feature checklist compared to the shorter exposure in the earlier study, both in the control condition and in the accurate composite condition. However, after being exposed to a misleading composite, there were fewer correct details and more incorrect details in the short (50 s) compared to the long target exposure (140 s).

In Exp. 2, participants answered the questions about the facial features and performed the lineup task after being exposed to the

TABLE 3 Effects of exposure to misleading composites on identification outcomes in target-present lineups

Authors (year)	N	Condition	Hit	Filler ID	False rejection	Filler ID or false rejection	Mix-up
Jenkins & Davies (1985)	144	CG/accurate	39.7	X	X	60.3	0.0
		Misleading	12.0	X	X	47.0	41.0
Davies & Jenkins (1985, Exp. 1) ^a	X	CG/accurate	57.5	41.5	X	X	X
		Misleading	39.5	32.5	X	X	25.5
Davies & Jenkins (1985, Exp. 2) ^a	X	CG/accurate	32.5	61.0	X	X	6.5
		Misleading	22.0	47.5	X	X	30.0
Gibling & Davies (1988)	194	CG/accurate	48.2	X	X	51.8	0.0
		Misleading	33.3	X	X	43.8	22.9
Franzen & Sporer (1994a)	79	CG/accurate	38.9	X	X	59.2	1.9
		Misleading	56.0	X	X	32.0	12.0
Franzen & Sporer (1994b)	89	CG/accurate	49.2	X	X	45.7	5.1
		Misleading	50.0	X	X	36.7	13.3
Dekle (2006)							
Lineups excluding do not knows	62	CG/accurate	45.0	X	X	55.0	X
		Biased	39.0	X	X	61.0	X
Showups excluding do not knows	47	CG/accurate	55.5	X	X	44.5	X
		Biased	42.0	X	X	58.0	X
Wells et al. (2005, Exp. 1)	100	No composite	84.0	6.0	10.0	16.0	X
		Composite	44.0	6.0	50.0	56.0	X
Kempen (2009, unpublished thesis)	81	No composite	65.1	11.6	23.3	34.9	X
		Composite	36.8	10.5	52.6	63.2	X
Unweighted means	99.5	CG/accurate	51.6	30.0	16.6	45.9	2.7
		Misleading	37.5	24.1	51.3	49.7	24.1
Total N	940						

Note: X = not possible, not reported, or impossible to reconstruct from reported data. CG: No Composite Control group; Accurate: Accurate Composite group; Misleading: One or several foils resemble misleading composite. Mix-up = Innocent suspect ID of a lineup member who was placed in the lineup due to his/her resemblance to the (misleading) composite.

^aPercentages were measured with a graphic program from scanned images of figs. 2 and 4 in the original publications. Therefore, percentages may not add up to 100% due to rounding or measurement imprecision.

misleading composite, a good composite or no composite. A facial feature checklist either preceded (prior recall) or followed composite exposure. As in Exp. 1, more correct and fewer incorrect details were recalled in the control and accurate composite conditions. In the misleading composite conditions, there were more correct answers and fewer incorrect answers after prior recall of features. It appears that the prior recall may have inoculated witnesses against the negative influence of a bad composite. Relatedly, a protective effect of re-reading one's own person descriptions was recently observed by Sporer, Davids, Kaminski, and McQuiston (2015), both after short and long testing intervals of 2 or 5 weeks.

However, Davies and Jenkins (1985) warn us about a potential problem in the procedure they used: If the descriptions do not contain important features or contain incorrect features, their protective function may be annihilated (see also the potentially negative effects of verbal overshadowing on face recognition [Alogna et al., 2014; Meissner & Brigham, 2001], discussion of which is beyond the scope of this paper).

Gibling (formerly Jenkins) and Davies (1988) also sought to examine if the "Guided Memory Interview" suggested by Malpass and Devine (1981), along with visual context reinstatement cues providing slides of the scene of the crime, would protect witnesses from a misinformation effect. Altering the hairstyle or adding a moustache to the composite reduced accuracy of recall compared to the control and accurate composite conditions, which showed no differences from each other. Although context reinstatement did reduce the misinformation effect, it did not fully eliminate the effects of a misleading composite.

5.2 | Summary of effects on face recall

In sum, studies that examined the effect of viewing a composite on recall of the perpetrator's face showed that exposure to a misleading composite impaired recall of facial features, whereas exposure to a good composite did not affect recall, as compared to viewing no composite at all. The harmful effects of viewing a misleading composite

TABLE 4 Effects of exposure to misleading composites on identification outcomes in target-absent lineups

Authors (year)	N	Condition	Correct rejection	Filler ID	Mix-up
Franzen & Sporer (1994a)	75	CG/accurate	48.0	50.2	2.0
		Misleading	52.0	32.0	16.0
Franzen & Sporer (1994b)	89	CG/accurate	64.4	23.7	11.9
		Misleading	40.0	20.0	40.0
Dekle (2006)					
Lineups excluding do not knows	63	CG/accurate	49.0	56.0	X
		Misleading	50.0	50.0	X
Showups excluding do not knows	98	CG/accurate	90.0	10.0	19.0
		Biased	85.0	15.0	9.0
Unweighted means		CG/accurate	62.9	35.0	11.0
		Misleading	56.8	29.3	21.7
Total N	325				
Average sample size	81.25				

Note: X = not possible, not reported, or impossible to reconstruct from reported data. CG: No Composite Control group; Accurate: Accurate Composite group; Misleading: One or several foils resemble misleading composite. Mix-up = Innocent suspect ID of a lineup member who was placed in the lineup due to his/her resemblance to the (misleading) composite.

were reduced by asking witnesses to recall the facial features of the perpetrator soon after the initial observation, or by helping them to reinstate the context of the event, before showing them the composite.

5.3 | Effects on face identification

Whereas the studies summarized thus far involved a change in modality, that is, analyzing misleading visual information and its effects on a verbal memory task, we now review studies that used a visual lineup task. The studies by Jenkins (Gibling) and Davies described above used both a verbal and a visual memory task (12-person identification lineups). Thus, the results of recall and recognition are not independent of each other.

We first present results for TP lineups (Table 3), followed by results for TA lineups (Table 4). The studies by Jenkins and Davies (1985), Davies and Jenkins (1985, Exp. 1 and 2), and Gibling and Davies (1988) described earlier did not report results separately for the no-composite and accurate composite groups because they found no differences between these conditions. While the studies by Franzen and Sporer (1994a, 1994b) and Dekle (2006) did report results separately for these two conditions, there were also no differences between them. Therefore, to make results comparable across studies, Tables 3 and 4 only report the percentages averaged across the no composite and the accurate composite groups and compare them to the misleading composite group. The studies by Wells et al. (2005, Exp. 1)⁶ and Kempen (2009) did not include a “good” composite condition.⁷

In the study by Jenkins and Davies (1985), lineup identification accuracy was rather poor. The misleading composite was associated with a strong reduction in hits and a strong increase in false

identifications of the misleading foil (mix-ups), compared to the combined no composite/accurate composite conditions. While effects on recall of incorrect features described above were stronger after longer retention intervals, the effects on identification were comparable in size for the three retention intervals.

In both experiments by Davies and Jenkins (1985), there were approximately⁸ 10–18% more hits and about 10% more filler identifications in the no/good composite conditions than in the misleading composite condition. Moreover, when a target-present lineup contained foils who displayed these misleading features, these foils were chosen almost as frequently as the real target observed (Davies & Jenkins, 1985, Exp. 1 and 2). In Experiment 1, the effects were stronger after a short target exposure, compared to a longer exposure. In Experiment 2, the composite exposure effect was much smaller when participants had answered a questionnaire on facial features before being exposed to the misleading composite. Taken together, it seems that the prolonged exposure in Experiment 1 and the prior recall in Experiment 2 protected witnesses' memories from the effect of the misleading composite (see also Sporer et al., 2015).

In the study by Gibling and Davies (1988), exposure to a misleading composite decreased hits by 15% and increased false identifications of designated misleading foils (mix-ups) by 23%, compared to the combined no composite/accurate composite conditions. The effects were much smaller when witnesses were interviewed with a “Guided Memory Interview” and slides from the scene which seemed to reinstate their memories both verbally and visually. This context reinstatement procedure did not only decrease the overall error rate, but also made it less likely that errors were induced by the misleading composite.

In the studies by Franzen and Sporer (1994a, 1994b), hit rates were almost identical in the no composite and good composite conditions. Whereas in the first study (Franzen & Sporer, 1994a), hit rates

were surprisingly (but not significantly) 17% higher in the misleading composite condition compared to the other two conditions, the means were virtually identical in the second study (Franzen & Sporer, 1994b). In both studies, there were slightly more mix-ups in the misleading composite condition compared to the other two conditions.

In a large-scale study (Dekle, 2006), participants viewed a slide show of a staged theft and provided a written description of the target. One day later, they were either (a) not shown a composite (control condition), (b) shown an “unbiased” composite, or (c) shown a “biased” composite. Two composites were used, both constructed by a police detective. Composite 1 was based on the “average” description of the perpetrator in a mock crime. For composite 2, a randomly selected participant provided direct feedback about the similarity to the police detective while he constructed the composite. In the “biased” condition, similarity ratings indicated that composite 1 from the average description was most similar to a foil, not the perpetrator. In the “unbiased” condition, all members of the photospread “were about equally similar” to composite 2 (see Dekle, 2006, p. 387).

Another day later, participants completed a free-recall task and one of five identification tasks: (a) a six-person TP lineup, (b) a six-person TA lineup, (c) a one-person TP showup, (d) a one-person TA showup, or (e) a one-person TA biased showup. We only report results here after excluding “Do not know” responses. Unexpectedly, Dekle found that viewing a composite, whether biased or unbiased, did not significantly affect identification accuracy in TP or TA lineups, despite the high statistical power of the experiment.

Most literature reviews have not discussed the earlier studies described here, instead focusing on a study by Wells et al. (2005), also with high statistical power, which yoked participants who constructed a composite to participants who were each exposed to a composite created by a member of the first group. Participants viewed a photo of a face for 180 seconds, while rating it on 10 different personality traits (e.g., attractive, intelligent, ...). Trait judgments have often been used to induce “deeper” encoding of faces, and are typically associated with better recognition (Sporer, 1991) and the production of better composites (Wells & Hryciw, 1984). Immediately after viewing the face, participants were asked to provide a verbal description of it. Next, participants in the Control condition went home, whereas participants in the View condition viewed a composite that had been constructed by another participant. Two days later, participants came back to the lab, received unbiased lineup instructions, and viewed a TP lineup consisting of six photos. Participants who had viewed a composite were significantly less likely to make a correct identification (44%) compared to the Control condition (84%). As Wells et al. emphasize, the strength of the study is the use of multiple stimulus faces from a large database, which compares favorably to other studies in terms of stimulus sampling (Wells & Windschitl, 1999). Unfortunately, there was no TA lineup for the composite viewing effect, and no data were reported on designated suspect misidentifications (mix-ups).

In an unpublished attempt to conceptually replicate the Wells et al. (2005) study, Kempen (2009, Experiment 1) introduced a few

methodological changes, in particular a shorter exposure time (5 s) and a change of pose of the face from frontal to three-quarter to reduce potential ceiling effects (in the control condition).⁹ Kempen used six different targets to prevent stimulus sampling issues and extensively piloted lineup fairness. Like Wells et al. (2005, Exp. 1), Kempen (2009) found that participants who had viewed a composite were significantly less likely to correctly identify the target in TP lineups (37% compared to 65% in the Control condition).

5.4 | Summary of hit rates in TP lineups

The sample sizes for the studies summarized in Tables 1 and 2 appear large enough to have sufficient power to detect a composite exposure effect on hits. To sum up, the earlier studies by Davies and colleagues (the top four studies in Table 2) reported small reductions in hits as a function of viewing misleading composites. The studies by Wells et al. (2005) and Kempen (2009) reported more substantial reductions in hits, while the studies by Franzen and Sporer (1994b) and by Dekle (2006) reported no effects, or only small effects, and the study by Franzen and Sporer (1994a) reported a nonsignificant tendency for higher hit rates in the misleading composite condition, contrary to expectation. Averaging across all 10 experiments, exposing participants to a misleading composite led to a 14% reduction in hits on TP lineups.

5.5 | Summary of filler IDs and false rejections in TP lineups

Because the articles reviewed did not consistently report comparable outcomes, no clear conclusions can be drawn regarding filler identifications and false rejections (see Table 2). While the two experiments by Davies and Jenkins (1985) reported a small drop in filler IDs as a result of viewing a misleading composite, the studies by Wells et al. (2005) and Kempen (2009) reported very few filler IDs, with no differences between the no composite and composite conditions. However, the latter two studies revealed a large increase in false rejections. Across all eight studies, there was only a small increase of about 4% in combined filler IDs and false rejections because of being exposed to a misleading composite.

5.6 | Summary of mix-ups in TP and TA lineups

In our view, the most crucial question is whether a misleading composite leads to a false identification of an innocent lineup member who resembles the composite more than the other lineup members. We refer to this type of false identification as a “mix-up,” that is the identification of a lineup member who was placed in the lineup due to his/her resemblance to the composite created by another witness. In TP lineups, mix-ups imply that the memory for the composite is stronger, or even replaces or “overwrites” the memory of the perpetrator. In TA lineups, mix-ups imply that the more recent memory of the

composite was strong enough for a witness to choose the suspect that the police had placed in the lineup due to his/her similarity to the composite created by another witness.

As the studies by Jenkins (Gibling) and Davies show, the effects of falsely identifying a foil displaying incorrect features like a moustache or curly (instead of straight) hair were substantial, particularly when encoding time was short and when memory was not protected via a prior recall task or a "Guided Memory Interview" or other context reinstatement procedures.

Surprisingly, only these earlier studies (the first four studies in Table 3), and the study by Dekle (2006, "showup condition") investigated whether the misleading composite led to a mix-up, while the other studies reported no data on this crucial question. Perhaps even more surprisingly, only three studies employed TA lineups (see Table 4). Only one of them showed a reduction in correct rejections in the misleading composite condition, together with an increase in mix-ups (Franzen & Sporer, 1994b).

A closer inspection of these studies suggests that mix-ups were moderated by other variables like exposure time, the timing of the misleading information, and lineup similarity. Mix-ups occurred more frequently after a short compared to a longer exposure time in Experiment 1 of Davies and Jenkins (1985). In Experiment 2, mix-ups were reduced when witnesses described the target's facial features prior to being exposed to the composite. Using a high-similarity lineup, in which all lineup members were selected in several pilot studies to resemble the target, also reduced mix-ups, compared to a more heterogeneous lineup, in which fillers were less similar to the target (Franzen & Sporer, 1994b; Sporer, 1996).

5.7 | Summary of effects on face identification

Taken together, the studies reviewed show that viewing a misleading composite can impair identification accuracy, but only after a short exposure to the target, if exposure to the composite is shortly before the memory test, and with a low-similarity lineup. Unfortunately, very few studies reported data on lineup similarity, and if they did, definitions of similarity differed widely across studies, sometimes operationalized as fillers matching a general description of the target, other times as ratings of target-filler similarity. Also, very few studies reported measures of lineup fairness.

We were surprised to find that most studies did not employ TA lineups, which are necessary to study the types of false identification that could have contributed (among other factors) to miscarriages of justice in real cases.

5.8 | Confidence-accuracy relationship in lineup identification

Exposure to a good or bad match composite may not only affect witness's B accuracy on a lineup identification task but also their

metacognitive processes and strategies, in particular their pre- and post-decision confidence and response latencies (see Franzen & Sporer, 1994a, 1994b, below). If witness B perceives A's composite to be a good likeness of the perpetrator, thus confirming her or his own memory, the witness's confidence to be able to identify the perpetrator in a lineup task is likely to be higher than if the composite is perceived as a rather poor likeness. Alternatively, one could hypothesize that when viewing a composite (of any quality likeness to the original perpetrator), a witness may retrieve their own memory of the perpetrator for comparison, which in turn may alter (increase or decrease) the witness's confidence in the memory image retrieved.

Relevant data were presented by Wells et al. (2005, Exp. 1) where the point-biserial correlation between confidence and accuracy was .12 in the composite viewing condition (which corresponds to Cohen's $d = .24$), compared to .59 ($d = 1.46$)¹⁰ in the no-composite control condition.¹¹ Thus, there appears to be a dramatic drop in the postdictive value of confidence when witnesses are exposed to a composite. Note that the confidence-accuracy relationship in the Wells et al. (2005) study included *all* decisions on TP lineups. In real crime cases, however, the decisions of people who choose someone from the lineup have a much greater potential impact than decisions of people who do not choose someone. Also, investigators and triers of fact do not know if the suspect is the criminal observed. Therefore, since the 1990s, confidence-accuracy relationships have usually been reported separately for choosers and nonchoosers. In the meta-analysis of 30 studies by Sporer, Penrod, Read, and Cutler (1995), confidence-accuracy relationships were generally much higher for choosers ($r_{pb} = .37$) than for nonchoosers ($r_{pb} = .12$), a finding replicated many times since (see also Wixted & Wells, 2017, for a new look at the importance of confidence to assess the accuracy of identifications). The effects of viewing a composite on the confidence-accuracy relationship should be further explored using confidence-accuracy-characteristic (CAC) curves as proposed by Mickes (2015).

The studies by Franzen and Sporer (1994a, 1994b) did report point-biserial correlations separately for choosers and nonchoosers. In the 1994a study (which had a total of 92 choosers), the point-biserial correlation was significant in the control condition ($r_{pb} = .42$), but not significant in the good composite ($r_{pb} = .19$) and bad composite conditions ($r_{pb} = .12$). In the follow-up study by Franzen and Sporer (1994b) with a total of 102 choosers, the correlation was significant in the good composite condition ($r_{pb} = .40$), but not significant in the bad composite ($r_{pb} = .17$) and no composite conditions ($r_{pb} = .00$). Unfortunately, the sample sizes were too small to calculate confidence-accuracy characteristic curves for these two studies. Further, witnesses were asked to rate their confidence both before and after conducting the lineup task. Recent research has shown that asking witnesses about their confidence before an identification task regarding whether they will be able to identify the perpetrator (i.e., pre-decision confidence) may reduce the diagnosticity of the post-decision confidence-accuracy relationship (Whittington et al., 2019).

6 | DISCUSSION

6.1 | Conclusions and implications

We critically examined studies which investigated whether viewing somebody else's composite (with a "good" or "bad" likeness to the perpetrator) affects later recall of facial features and the identification of the perpetrator in a lineup. The studies did not provide sufficient data and/or were too heterogeneous methodologically to allow a formal meta-analytic integration. Hence, we opted for a thorough, detailed critical narrative review. Based on the evidence we conclude that viewing someone else's composite, particularly if it is misleading, may harm subsequent memory for the perpetrator, affecting the recall of specific features as well as identification accuracy.

Some studies showed that viewing a misleading composite led to a reduction of hits in TP lineups. The implication is that the police may lose incriminating evidence against a suspect when a witness has observed another witness's (incorrect) composite. Note that some of the misleading composite manipulations were rather strong, such as altering hairstyle and adding a moustache. These features play an important part in recognizing a face, especially after a significant delay between encoding and exposure to misinformation (Sporer & Horry, 2011). It appears that only substantially misleading composites have a detrimental effect on hits, while exposure to "good" or "accurate" composites yield results that are no different from those of no composite control groups.

Some studies also reported filler identifications, but because filler identifications in both TP and TA lineups are "known errors", filler identifications do not answer the question whether misleading composites lead to an increase in innocent suspect identifications, and thus ultimately to wrongful convictions. Only studies that assess the effects of a misleading composite on designated suspect identifications (i.e., a lineup member resembling the composite) in TA lineups can answer this question directly. Unfortunately, of the 10 experiments reviewed, only three reported designated suspect identifications (Dekle, 2006; Franzen & Sporer, 1994a, 1994b; see Table 4).

On the positive side, it is comforting to know that available evidence suggests that harmful effects of viewing composites are short-lived, possibly limited to cases when the misleading composite was introduced shortly before the lineup task (rather than immediately after viewing the target), and can be reduced in magnitude or even reversed (Davies & Jenkins, 1985; Gibling & Davies, 1988; Jenkins & Davies, 1985). In particular, from a system variable perspective it is good to know that context reinstatement procedures like a "Guided Memory Interview" and providing visual context cues as well as constructing fair, high similarity ("homogeneous") lineups may reduce composite misinformation effects (Franzen & Sporer, 1994a, 1994b; Gibling & Davies, 1988; Sporer, 1996).

The emphasis on a high similarity lineup squares well with recent policy recommendations emphasizing the use of fair lineups, which may create a "filler-siphoning effect", spreading filler identifications across lineup members (Wells et al., 2020). In the study by Franzen and Sporer (1994b), mix-ups were nearly halved in high similarity

lineups, and filler identifications were spread across all fillers (see figs. 2 and 3 in Sporer, 1996).

6.2 | Guidelines for practice

We note a few specific practical recommendations for police and criminal justice officials who deal with witnesses who have been exposed to face composites. In training, investigating officers, attorneys and judges should be made aware of potential positive or negative effects of publishing composites.

Composites are usually published to draw attention to a case and to recruit assistance from the public to help the police find suspects. Police and fact-finders should keep verbatim records (of audio- or videotapes) of interviews with such witnesses or members of the public that they interview in connection with the case. This is especially important when they are interviewing witnesses to the original crime who did not create the original composite. They should ask questions of such witnesses to create a *timeline of events*: (a) When and where did the witness read about the case and see the composite picture? (b) Did the witness save a copy and how often did they look at it? When was the last time they viewed the composite? (c) Did the witness talk to anybody about it and/or show it to anybody else? (d) How good a likeness of the perpetrator does the witness think the composite is? (e) Which parts of the perpetrator's face are portrayed well, and which parts poorly? (f) How confident is the witness that they could identify the perpetrator in a lineup? In addition to creating a timeline of events, the police should take potential effects of exposure to the composite into account when assessing lineup fairness. For example, lineups could be shown to mock witnesses with or without the composite. Admittedly, these guidelines are tentative and should be tested empirically. In the following we sketch some gaps in knowledge and outline what we think future studies should investigate.

6.3 | What would an ideal study look like?

Most studies reported here were conducted with older, feature-based composite systems. Clearly, similar studies with newer, holistic composite construction systems (e.g., Hancock, 2000; Saraiva et al., 2018; Skelton et al., 2019; Tredoux et al., 1999; Tredoux et al., 2006) are necessary to update our knowledge. Because witnesses may be repeatedly exposed to composites constructed by others, for example via television, the press, and social media, future studies should not only use different retention intervals, but also investigate repeated exposure. In real cases, witnesses who are asked to attend a lineup procedure are likely to "refresh" their memories by looking at past descriptions and composites published in newspapers or other media.

An important avenue for further research that to our knowledge has not yet been addressed in the literature is how an identification decision of witness B, who has not created a composite, may be affected by her or his appraisal of witness's A composite.¹² We can

only speculate on potential metacognitive processes that might be evoked. If witness B considers A's composite a badly created likeness in the sense that it diverges in major ways from her or his memory ("His nose was much bigger, and his face much rounder"), the composite may simply be ignored. Similarly, if the composite contained a distinctive feature (like a large moustache that witness B is confident the perpetrator did not have), she or he will also discard this new, discrepant information (see the study by Strack & Bless, 1994, who examined the metacognitive and presuppositional strategies witnesses may engage in with discrepant verbal misinformation). On the other hand, if witness B considers A's composite a good likeness, they may use this information to reinforce their memory. In the latter case, if A's composite does indeed resemble the perpetrator well, the chances of a correct identification should increase. On the other hand, if the memory of witness B has faded (due to a long delay), witness B may cling to the information contained in A's composite—even if it is partially wrong—and may consequently misidentify somebody resembling this composite (as in the mix-ups in Franzen & Sporer, 1994a, 1994b) studies.

Studies should carefully manipulate and pilot-test the similarity of composites to designated suspects in TA lineups, and analyze mix-ups, rather than simply reporting hits and filler identifications in TP lineups. In line with current identification practice recommendations, post-decision confidence should be assessed immediately after an identification decision, and the whole procedure should be videotaped (Sporer, 1992b; Wells et al., 2019). Analyses should focus on responses by choosers, and newer confidence calibration analyses and CAC curves could also be considered for analysis (Brewer, Weber, & Guerin, 2019; Mickes, 2015; Wixted & Wells, 2017). It is an intriguing empirical question whether or not exposure to a (misleading) composite affects the probative value of confidence and decision times (Franzen & Sporer, 1994a, 1994b; Wells et al., 2005).

An additional open question is whether the use of a Cognitive Interview would produce better composites analogous to improved person descriptions, which in turn could reduce the effects of "misleading" composites (but see Koehn & Fisher, 1997, who found no improvement in composite quality when using a "guided memory" technique). Finally, could warnings given before an identification task regarding the possibility that a composite viewed earlier may contain errors, reduce detrimental composite exposure effects (cf. Blank & Launay, 2014)? Perhaps interviewing witnesses before an identification task regarding the exposure(s) to composites is a good idea. The witness could be asked how well the composite matched their own memory of the perpetrator, and when and how often they saw the composite. In short, there are many issues regarding composite exposure effects still awaiting an answer.

ACKNOWLEDGEMENTS

This research was supported by a grant to Siegfried L. Sporer on "Recognizing Faces of Other Ethnic Groups" by the Deutsche Forschungsgemeinschaft (German Science Foundation: Grant Nr. Sp272/9-1). Collection of articles, data coding and preliminary analyses were supported by a joint grant to Colin G. Tredoux

and Siegfried L. Sporer on "Reducing False Identifications of Other-race People through new Computer Techniques" (SUA 07/028) by the National Research Foundation of South Africa and the Deutsche Zentrum für Luft- und Raumfahrt des Bundesministerium für Bildung und Forschung (German Ministry of Science and Education). Finally, the article was made possible with the support from the South African Centre for Digital Language Resources (SADiLaR). SADiLaR is a research infrastructure established by the Department of Science and Technology of the South African government as part of the South African Research Infrastructure Roadmap (SARIR).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Siegfried L. Sporer  <https://orcid.org/0000-0002-7236-4300>

Colin G. Tredoux  <https://orcid.org/0000-0002-9653-786X>

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

ENDNOTES

- ¹ These newer composite systems (e.g., EFit-V or EFit6, Evofit; ID) are based on statistical models and learning algorithms, typically presenting witnesses with an array of faces and asking them to select the face(s) in the array that have a likeness to the target. Based on the selection, new sets of faces are iteratively created in succeeding trials, shown in simulations to progressively approximate the true appearance of the target (Frowd, Hancock, & Carson, 2004; Gibson, Solomon, & Bejarano, 2003; Tredoux, Rosenthal, da Costa, & Nunez, 1999; see also Tredoux, Nunez, Oxtoby, & Prag, 2006, for reports of simulations). Importantly, no disembodied features are shown for selection, thus encouraging "holistic processing" of faces. However, it is possible that presenting many arrays of increasingly similar faces—as is the case for these systems—could interfere with the original memory trace.
- ² We conducted a meta-analysis to assess whether the construction of facial composites affects witnesses' subsequent lineup identification decisions. We located 23 studies ($k = 56$ effects, $N = 2,276$ participants). Using a random-effects model, we found no significant effects of composite construction on correct identifications from target-present lineups ($k = 23$), nor on incorrect identifications from target-absent lineups ($k = 12$). We found some weak evidence that composite construction *reduced* incorrect identifications from target-present lineups ($k = 21$). Because there was no significant heterogeneity in effects for any of the dependent variables (once outliers were removed), moderator analyses were not conducted.
- ³ Note that the term "bad composite" can have several different meanings: (a) a composite that has little resemblance to the target (as can be assessed via pilot studies in experimental studies); (b) a composite that has little resemblance to the target according to the witness creating it (as assessed by self-ratings), and (c) a composite that has little resemblance to the target, as rated from the perspective of a second witness who has also observed the target but not constructed a composite. The second witness's perception could also be assessed via ratings of the first witness's composite.

- ⁴ The results from Dekle (2006) using a lineup task versus a showup task are reported as separate “experiments” because outcomes were reported differently.
- ⁵ It is an open empirical question if comparable results would have been obtained if free recall or open-ended questions (cued recall) had been used instead of these feature lists.
- ⁶ Exp. 2 did not include a composite exposure group.
- ⁷ Note that in a pilot study for the Wells et al. experiment, the average matching rate to the target was only 25%, indicating that the composites were rather poor likenesses. Nonetheless, there was a positive correlation between hit rates in TP lineups and ratings of composite quality ($r_{pb} = .47$). Composite quality was also related to the likelihood of a correct identification in some other studies (e.g., Holland, Otzen, & Sporer, 1994).
- ⁸ Percentages were measured with a graphic program from scanned images of figs. 2 and 4 in the original publications.
- ⁹ The shorter encoding time was used after a previous attempt to replicate the Wells et al. (2005) procedure with the same 180 s exposure time to rate the faces resulted in ceiling effects of 100% hit rates in the no construction and 98% in the composite construction group (Maskow, Schmidt, Tredoux, & Nunez, 2007, Exp. 1). Correct rejection rates in TA lineups were also above 90%. In Exp. 2, which used a 16 s exposure time for the same rating task, hit rates were still above 80% and correct rejection rates above 60%. Based on these failures to replicate the Wells et al. procedure with long exposure times, Kempen (2009) chose a much shorter exposure time of 5 s.
- ¹⁰ Note that in the literature on the confidence-accuracy relationship, many authors have incorrectly described *point-biserial* correlations of .10, .24, and .37, respectively, as low to medium associations, apparently neglecting the distinction between point-biserial correlations and correlations between two continuous variables, for which .10, .30, and .50 are considered low, medium, and high associations. Following Cohen's (1988) rule-of-thumb recommendations, point-biserial values of .10, .24, and .37 correspond to Cohen's *d* values of 0.10, 0.50, and 0.80, respectively, which are usually labeled as small, medium, and large effect sizes.
- ¹¹ The relationship in the control condition appears very high for TP lineups and compared with studies reporting the relationship across the whole sample (e.g., an average $r_{pb} = .08$ in Wells & Murray, 1984, or $r_{pb} = .25$ in Sporer et al., 1995).
- ¹² We are indebted to an anonymous reviewer for raising this issue.

REFERENCES

- Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., ... Zwaan, R. A. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, 9, 556–578. <https://doi.org/10.1177/1745691614545653>
- Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, 3, 77–88. <https://doi.org/10.1016/j.jarmac.2014.03.005>
- Brewer, N., Weber, N., & Guerin, N. (2019). Police lineups of the future? *American Psychologist*, 75, 76–91. <https://doi.org/10.1037/amp0000465>
- Bruce, V., Ness, H., Hancock, P. J., Newman, C., & Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, 87(5), 894–902. <https://doi.org/10.1037/0021-9010.87.5.894>
- Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 739–755. <https://doi.org/10.1037/0278-7393.30.4.739>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- 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. (2007). Facial composites: Forensic utility and psychological research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Volume 2: Memory for people* (pp. 59–83). Mahwah, NJ: Lawrence Erlbaum Associates.
- Davis, D., & Loftus, E. F. (2007). Internal and external sources of misinformation in adult witness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol I: memory for events* (pp. 195–237). Mahwah: Lawrence Erlbaum Associates Publishers.
- Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, 30(3), 287–307. <https://doi.org/10.1007/s10979-006-9008-1>
- Dekle, D. J. (2006). Viewing composite sketches: Lineups and showups compared. *Applied Cognitive Psychology*, 20, 383–395. <https://doi.org/10.1002/acp.1185>
- Fodarella, C., Frowd, C. D., Warwick, K., Hepton, G., Stone, L. D., Date, L., & Heard, P. (2017). Adjusting the focus of attention: Helping witnesses to evolve a more identifiable composite. *Forensic Research & Criminology International Journal*, 5(1), 00143. <https://doi.org/10.15406/frcij.2017.05.00143>
- 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. <https://doi.org/10.1108/JFP-10-2014-0033>
- Franzen, S., & Sporer, S. L. (1994a). Personenverwechslungen durch irreführende Rekonstruktionsbilder: Zum Einfluss nachträglicher Informationen und der Wiederherstellung des Wahrnehmungskontextes [Person mix-ups as a function of misleading composites: On the influence of postevent information and context reinstatement]. In S. L. Sporer & D. Meurer (Eds.), *Die Beeinflussbarkeit von Zeugenaussagen [Influencing eyewitness testimony]* (pp. 207–236). Marburg, Germany: N.G. Elwert.
- Franzen, S., & Sporer, S. L. (1994b). Personenverwechslungen und Möglichkeiten ihrer Vermeidung: Können Augenzeugen durch Visualisierung gegen den Einfluss von irreführenden Rekonstruktionsbildern immunisiert werden [Person mix-ups and possible countermeasures: Can eyewitnesses be inoculated against misleading composites through visualization]? In S. L. Sporer & D. Meurer (Eds.), *Die Beeinflussbarkeit von Zeugenaussagen [Influencing eyewitness testimony]* (pp. 237–283). Marburg, Germany: N.G. Elwert.
- Frowd, C. D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, 10, 63–81. <https://doi.org/10.1348/135532504x15358>
- 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. <https://doi.org/10.1081/10683160310001634313>
- Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT. *ACM Transactions on Applied Perception*, 1(1), 19–39. <https://doi.org/10.1145/1008722.1008725>
- Frowd, C. D., Valentine, T., & Davis, J. P. (2015). Facial composites and techniques to improve image recognizability. In T. Valentine & J. P. Davis (Eds.), *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 43–70). Chichester, England: Wiley-Blackwell.

- Garrett, B. (2011). *Convicting the innocent*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/harvard.9780674060982>
- Gibling, F., & Davies, G. M. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology*, 79, 129–141. <https://doi.org/10.1111/j.2044-8295.1988.tb02278.x>
- Gibson, S. J., Solomon, C. J., & Bejarano, A. P. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. Paper presented at: Proceedings of the British Machine Vision Conference, Norwich, UK. doi:<https://doi.org/10.5244/c.17.23>
- Gronlund, S. C., & Carlson, C. A. (2014). System-based research on eyewitness identification. In T. J. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 595–613). London, England: Sage.
- Hancock, P. J. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments, & Computers*, 32(2), 327–333. <https://doi.org/10.3758/bf03207802>
- Hasel, L. E., & Wells, G. L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and Human Behavior*, 31, 193–207. <https://doi.org/10.1007/s10979-006-9007-2>
- Holland, K., Otzen, H., & Sporer, S. L. (1994). Der Einfluss der Rekonstruktion und der Beschreibung von Gesichtern auf das spätere Wiedererkennen [The influence of constructing a face composite and describing a face on later face recognition]. In S. L. Sporer, & D. Meurer (Eds.), *Die Beeinflussbarkeit von Zeugenaussagen [Influencing eyewitness testimony]* (pp. 154–206). Germany: N. G. Elwert.
- Jenkins, F., & Davies, G. M. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, 70, 164–176. <https://doi.org/10.1037/0021-9010.70.1.164>
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28. <https://doi.org/10.1037/0033-2909.114.1.3>
- Kempen, K. (2009). "It's the thought that counts": Face composite production can hamper recognition performance. (Unpublished honour's thesis). University of Cape Town.
- Koehn, C. E., & Fisher, R. P. (1997). Constructing facial composites with the Mac-a-Mug pro system. *Psychology, Crime & Law*, 3, 209–218. <https://doi.org/10.1080/10683169708410815>
- 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.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361–366. <https://doi.org/10.1101/lm.94705>
- Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior*, 4, 323–334. <https://doi.org/10.1007/BF01040624>
- Malpass, R. S., & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology*, 66, 343–350. <https://doi.org/10.1037/0021-9010.66.3.343>
- Maskow, M., Schmidt, H., Tredoux, C. G., & Nunez, D. T. (2007). Face composite production does not affect eyewitness identification accuracy. Unpublished manuscript, University of Cape Town.
- Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603–616. <https://doi.org/10.1002/acp.728>
- Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, 4, 93–102. <https://doi.org/10.1016/j.jarmac.2015.01.003>
- Saraiva, R. B., Castillo, G. M., Nogueira, R. N., Coelho, L. A. M., Alarcão, L. C. P., & Lage, J. (2018). Eyewitnesses memory for faces in actual criminal cases: An archival analysis of positive facial composites. *Estudos de Psicologia*, 22(3), 247–256.
- Schmidt, H. C., & Tredoux, C. G. (2006). Utilisation and usefulness of face composites in the South African police service: An evaluation study. *South African Journal of Criminal Justice*, 19, 303–314.
- Shepherd, J. W., & Ellis, H. D. (1996). Face recall—Methods and problems. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87–115). Hillsdale, NJ: Lawrence Erlbaum.
- Skelton, F. C., Frowd, C. D., Hancock, P. J. B., Jones, H. S., Jones, B. C., Fodarella, C., ... Logan, K. (2019). Constructing identifiable composite faces: The importance of cognitive alignment of interview and construction procedure. *Journal of Experimental Psychology: Applied*. <https://doi.org/10.1037/xap0000257>
- Snetkow, W. A., Sinin, M. A., & Delang, A. (1981). *Das subjective Porträt [the subjective portrait]*. Berlin, Germany: Ministerium des Innern.
- Sporer, S. L. (1991). Deep—Deeper—Deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 323–333. <https://doi.org/10.1037/0278-7393.17.2.323>
- Sporer, S. L. (1992a). Person descriptions in an archival analysis of criminal cases. Paper presented at the Biennial Meeting of the American Psychology-Law Society in San Diego, California.
- Sporer, S. L. (1992b). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*, 22(2), 157–180. <https://doi.org/10.1002/ejsp.2420220205>
- Sporer, S. L. (1996). Experimentally induced person mix-ups through media exposure and ways to avoid them. In G. M. Davies, S. Lloyd-Bostock, M. McMurrin, & C. Wilson (Eds.), *Psychology and law: Advances in research* (pp. 64–73). Berlin, Germany: De Gruyter.
- Sporer, S. L., Davids, M. C., Kaminski, K. S., & McQuiston, D. (2015). The verbal facilitation effect: Re-reading person descriptions as a system variable to improve identification performance. *Memory*, 24, 1329–1344. <https://doi.org/10.1080/09658211.2015.1106561>
- Sporer, S. L., & Horry, R. (2011). Pictorial versus structural representations of ingroup and outgroup faces. *Journal of Cognitive Psychology*, 23(8), 974–984. <https://doi.org/10.1080/20445911.2011.594434>
- Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315–327. <https://doi.org/10.1037/0033-2909.118.3.315>
- Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, 33, 203–217.
- Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: An eigenface based construction system. *South African Computer Journal*, 37, 90–97.
- Tredoux, C. G., Rosenthal, Y., da Costa, L., & Nunez, D. (1999, July). Face reconstruction using a configural, eigenface-based composite system. Paper presented at: The Meeting of the Society for Applied Research in Memory and Cognition (SARMAC), in Boulder, CO.
- Tredoux, C. G., Sporer, S. L., Vredeveltd, A., Kempen, K., & Nortje, A. (2020). Does constructing a facial composite affect eyewitness memory? A research synthesis and meta-analysis. *Journal of Experimental Criminology. Advance online publication. Advance online publication*. <https://doi.org/10.1007/s11292-020-09432-z>
- 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. <https://doi.org/10.1037/1076-898x.11.3.147>
- Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, 12(4), 338–344.
- Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*. 44 (1), 3–36. <http://dx.doi.org/10.1037/lhb0000359>.
- Wells, G. L., Murray D. M. (1984). Eyewitness confidence. In G. L. Wells, & E. F. Loftus, (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York: Cambridge University Press.

- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9), 1115–1125. <https://doi.org/10.1177/01461672992512005>
- Whittington, J. E., Carlson, C. A., Carlson, M. A., Weatherford, D. R., Krueger, L. E., & Jones, A. R. (2019). Asking an eyewitness to predict their later lineup performance could harm the confidence-accuracy relationship. *Applied Cognitive Psychology*, 34, 119–131. <https://doi.org/10.1002/acp.3599>
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological*

Science in the Public Interest, 18, 10–65. <https://doi.org/10.1177/1529100616686966>

How to cite this article: Sporer SL, Tredoux CG, Vredeveldt A, Kempen K, Nortje A. Does exposure to facial composites damage eyewitness memory? A comprehensive review. *Appl Cognit Psychol*. 2020;34:1166–1179. <https://doi.org/10.1002/acp.3705>