The impact of physical and mental reinstatement of context on the identifiability of facial composites

by

Cristina Fodarella

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire

July 2020

STUDENT DECLARATION FORM

Concurrent registration for two or more academic awards

*I declare that while registered for the research degree, I was with the University's specific permission, a registered candidate for the following award:

'Associate Fellow of The Higher Education Academy'

Material submitted for another award

*I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

*Not applicable

Signature of Candidate L. Fadiselle

Type of Award **Doctor of Philosophy**

School **Psychology**

Abstract

Numerous studies demonstrate that memory recall is improved by reflecting upon, or revisiting, the environment in which information to-be-recalled was encoded. The current thesis sought to apply these 'context reinstatement' (CR) techniques in an attempt to improve the effectiveness of facial composites—likenesses of perpetrators constructed by witnesses and victims of crime. Participant-constructors were shown an unfamiliar target face in an unfamiliar environment (e.g., an unknown café). The following day, participants either revisited the environment (physical context reinstatement) or recalled the environmental context in detail along with their psychological state at the time (mental context reinstatement, Detailed CR); they then freely recalled the face and constructed a facial composite using a holistic (EvoFIT) or a feature system (PRO-fit). Over the course of five experiments and meta-analyses, Detailed CR of the environmental context was effective at increasing correct naming and likeness ratings of ensuing composites. The size of the advantage for Detailed CR was dependent on the extent to which the environment had been encoded: the advantage was (i) variable for incidental encoding (Experiments 1-3) with an overall small effect size (ES) (assessed by meta-analysis), (ii) best (very large ES) under intentional encoding (Experiment 3) and (iii) intermediate (large ES) for incidental encoding when participants were encouraged to engage naturally with the environment (Experiment 4). Detailed CR was also found to be effective when combined with a specific interviewing technique (Holistic-Cognitive Interview) where constructors focused on the target's character; it was no more effective when constructors were prompted to recall the environment in greater detail. Further analyses (Meta-analyses) and additional data (Experiment 5) indicate that the advantage of Detailed CR was mediated by an increase in constructor's total face recall. Results are interpreted in terms of the encoding specificity principle and can be applied by forensic practitioners who use feature and recognition systems.. This thesis is the first to reveal that context cues can be implemented effectively during forensic face construction using modern composite systems.

Table of contents

List of tables and figures

CHAPTER 1

CHAPTER 2

APPENDIX 1

Acknowledgements

First of all, I would like to thank my Director of Studies, Charlie Frowd, for all of his expertise, support and guidance—not only during my Ph.D but also during the many, many years leading up to it. If it wasn't for him, I don't think I would be where I am now—not least because he has sparked my excitement for facial composites during my 2nd year of undergraduate studies. He has become not only a great colleague and role model to me, but also a remarkable friend. Thank you!

I am also deeply grateful for all of the advice, insight and collaboration from the rest of my supervisory team: John Marsh and Simon Chu as well as (my 'old' team) Palwinder Athwal-Kooner and Rachel Wilcock. A special thanks to John who has been a great friend to me throughout, always caring and encouraging.

I would also like to thank Elizabeth Jackson and Ellena Wood for having collected some excellent data for my thesis. Thank-you also to Faye Skelton and Helen Jones for their invaluable input into improving the research paper that is now Chapter 2.

UCLan has always been an amazing place to have both studied and worked in, and deserves a special acknowledgement. It is not only full of lovely colleagues, but I have also made many friends along the way. A special thank-you to Rachel Thorley (Dr Joyce!), whose Ph.D journey started and ended about the same time as my own: you've become a close friend to me—thanks for being you!

Not many people would be able to complete a Ph.D thesis without the support from friends and family outside of academia—I would like to thank all of mine (too little space to name all!). A special thanks goes to my sister Angela, who is one of the most important people in my life: a wonderful person and my best friend. I am also deeply grateful to my partner (and best friend) Ben, who has not only helped by proof-reading my chapters and acting as a second coder for some of my stats, but also by keeping me sane, loved and happy during the process (…and well fed!). I'd like to thank my aunt Andrea and uncle Franco for their on-going love, interest, and support, as well as those amazing Spa trips that helped me re-energise—thank you!

Last but not least, I owe special thanks to Lady P who—by purring and falling asleep on my lap—has helped me be more productive during my many, many hours of writing.

In loving memory of Mama, Papa and Nonna.

Abbreviations

ASD	Autism Spectrum Disorder
CI	Cognitive Interview
CR	Context Reinstatement
DV	Dependent Variable
ES	Effect size
FM	Focused meditation
GEE	Generalised Estimating Equations
H-CI	Holistic-Cognitive Interview
ICE	Item, context and ensemble information
IQR	Inter-quartile range
M	Mean
MCR	Mental Context Reinstatement
MD	Mean difference
PCR	Physical Context Reinstatement
SD	Standard Deviation
SE	Standard Error
TAP	Transfer-appropriate processing
TV	Television
UCLan	University of Central Lancashire
UK	United Kingdom
UOI	Unit of information
US	United States

1

LITERATURE REVIEW

General introduction

The current chapter will first provide an overview of the literature and theories regarding context reinstatement (CR) effects; that is, a technique to facilitate memory recall which uses the physical or a mentally-visualised environment, that was present during time of encoding. The review will focus more-specifically on CR effects on verbal, eyewitness and facial memory. CR effects on facial memory will be reviewed more-intensely as it is more-closely related to research of the current thesis. Since the CR technique can improve eyewitness and facial memory, the thesis aims to utilise it to facilitate facial composite production. In a forensic setting, facial composites are facial likenesses of a perpetrator constructed from an eyewitness memory. Following on from a review of the CR literature, an overview of facial composite systems will be outlined. Existing techniques of how to improve composite effectiveness will be discussed, before explaining how CR could be applied prior to composite construction.

Introduction to context reinstatement effects

Context reinstatement (CR) effects refer to the notion that reinstating the original context that was present during time of encoding can trigger or enhance memory of the to-be-recalled target memory. Context does not only have to be the actual environment (e.g., scene of the crime) but could be the psychological and physical state of the observer as well as any other visual cues (e.g., clothing of persons present) available during the time of encoding. The encoding specificity principle (Tulving & Thomson, 1973) is used to explain this phenomenon. It stipulates that a new, target memory is encoded alongside a network of any other information available during the time of encoding. This creates an associative network of information that form retrieval cues to one another. These cues can enable access to the target memory (e.g., clothing or the appearance of a face) during retrieval.

The empirical literature provides support for the hypothesis that context cues can facilitate access to a desired memory, revealing positive CR effects by greater memory retrieval (e.g., see Smith, 2013 for a review; see Smith & Vela, 2001 for a review and meta-analysis). The first experimental study known to the author that has investigated CR effects occurred a century ago when Carr (1917) manipulated a maze-running environment for rats: maze memory was found to be better when lighting matched during learning and test. Following on from animal studies (see Smith, 2013 for an overview of CR tests with animals), research has focused on human participants, and investigated CR effects for verbal memory (e.g., Campeanu, Craik, Backer, & Alain, 2014; Godden & Baddeley, 1975; Smith, 1979), eyewitness memory (e.g., Dando, 2013; Wagstaff, Wheatcroft, Caddick, Kirby, & Lamont, 2011; Wong & Read, 2011), facial memory (e.g., Davies & Milne, 1985; Shapiro & Penrod, 1986) as well as object memory (Barak, Vakil, & Levy, 2013; Koen, Aly, Wang, & Yonelinas, 2013; Levy, Rabinyan, & Vakil, 2008). The former three topics are reviewed below (please note that object memory has only been scarcely researched relative to the above-outlined stimuli, and is less relevant to the research of the current thesis).

Verbal memory

In verbal learning, Godden and Baddeley (1975) conducted an experiment, now frequently cited in this context. A group of divers encoded a word list either on land or under water, recalling words either in the same or different environment. Returning to the original environment led to a significant increase in number of words recalled compared to a different environment. This finding of congruency was later supported by Smith and Sinha's (1987) study who also used a distinctive environment, a flotation tank. Ensuring that the two environments are distinctive and perceptually different to one another is more likely to promote a CR effect as perceptually rich stimuli can act as recall cues. That said, some studies have found a CR effect simply by using two different types of classrooms, with recall being greater when rooms for testing and recall are the same (Fernández & Alonso, 2001; Fernández & Glenberg, 1985; Metzger, Boschee, Haugen, & Schnobrich, 1979; Smith, 1979; Smith, Glenberg, & Bjork, 1978).

Memory recall is not only facilitated by physically reinstating the environmental context as outlined above, but also by mentally visualising it (e.g., Smith, 1979; a more detailed description of this Mental Context Reinstatement is incorporated later in this review). In line with this finding, there is a technique of how to overcome detrimental effects of environment change from learning to testing: mentally pre-instating the (future) testing environment helps to avoid negative effects from context change (Brinegar, Lehman, & Malmberg, 2013; Chu, Handley, & Cooper, 2003). Research on transfer-appropriate processing (TAP) is also relevant to CR effects: TAP refers to the notion that there is increased memory retrieval when cognitive processes during learning and testing are matched (e.g., see Karpicke, Lehman, & Aue, 2014). This research area can therefore be seen as an extension of context effects.

However, context does not need to be the physical environment (e.g., Godden & Baddeley, 1975, 1980; Greenspoon & Raynard, 1957; Hockley, 2008; Smith et al., 1978), it can also be represented in a variety of different ways: as background images in a video clip (Smith & Manzano, 2010), the paired word that accompanies a to-be-remembered word (Meyer & Schvaneveldt, 1976; Tulving & Thomson, 1971), the position of a word on the screen (Macken, 2002; Murnane & Phelps, 1993), the sentence in which a target word is embedded (Fisher & Craik, 1977, 1980; Kotchoubey & El-Khoury, 2014; Tulving & Gold, 1963), a word in which the to-be-remembered letter is embedded (e.g., Reichler, 1969; Wheeler, 1970), a background colour of target words (Dulsky, 1935), the voice in which a word was spoken (e.g., Bradlow, Nygaard, & Pisoni, 1999; Campeanu et al., 2014; Craik & Kirsner, 1974; see Campeanu, Craik, & Alain, 2015 for a review), the audience who are present during learning (Burri, 1931), and odours present during testing (Ball, Shoker, & Miles, 2010). These contexts have repeatedly led to positive contextual effects, with a significantly greater number of words remembered compared to controls, effects that are evident not only for adults but also for children (Jensen, Harris, & Anderson, 1971).

Further retrieval cues that successfully aid verbal memory recall include psychological states. Memory recall is facilitated by reproducing participants' mood (Bower, 1981; Bower & Cohen, 1982; Bower, Monteiro, & Gilligan, 1978; Eich & Metcalfe, 1989) or level of arousal (Clark, Milberg, & Ross, 1983) during testing. There is also evidence for state-dependent learning: memory recall is improved when the pharmacological states, such as sleep deprivation or the influence of drugs or alcohol, are reinstated (see Eich, 1980; Eich, Weingartner, Stillman, & Gillin, 1975).

Despite the fact that the majority of studies have shown that CR effects facilitate memory (all of the above), some studies have failed to support an increase in recall (Cousins & Hanley, 1996; Eich, 1985; Farnsworth, 1934; Fernández & Alonso, 2001; Fernández & Glenberg, 1985; Reed, 1931; Saufley, Otaka, & Bavaresco, 1985), or even found reduced performance (Wilhite, 1991), thereby indicating that there remain theoretical and conceptual issues unresolved in the context memory literature. It is conceivable that some of these inconsistencies could be explained by the Outshining hypothesis (Smith, 1988, 1994; Smith & Vela, 1986; see Smith, 2013), Cue overload/fan effect (e.g., Watkins & Watkins, 1975) and Overshadowing hypothesis (see Matzel, Schachtman, & Miller, 1985; Smith, 2013; Smith & Vela, 2001). These hypotheses may mediate the extent to which context is utilised as a retrieval cue and are outlined in more detail below.

Outshining hypothesis

The outshining hypothesis (Smith, 1988, 1994; Smith & Vela, 1986; see Smith, 2013) postulates that, depending on the test conditions, environmental information may not actually be used by the participants to aid their memory. More-specifically, the extent to which non-contextual memory cues are presented during testing reduces or even eliminates reliance on contextual cues. Cued recall conditions provide certain cues to facilitate memory which may in turn reduce the dependence on contextual cues. Recognition tests take this notion even further and provide the to-be-remembered item, thereby further-reducing or even eliminating—'outshining'—the need for contextual cues altogether.

Cue overload/fan effect

The cue overload effect (e.g., Watkins & Watkins, 1975) or fan effect (Anderson, 1974) is related to the outshining hypothesis. The former is based on the concept that context cues are more likely to facilitate a target memory when there are fewer competing targets. In other words, when context cues are 'overloaded' with memory targets, it is increasingly difficult or less likely to gain access to a target memory. The fan effect (Anderson, 1974) similarly indicates that response times or error rates increase when there is more information associated with the target memory. Therefore, if the target memory has many additional, competing targets (or if the size of the context cue's fan is smaller), context cues are less likely to be used if other recall cues are also presented, thereby demonstrating an outshining effect. If there are fewer targets, it is thus more likely that a CR effect will be observed.

Empirical evidence has emerged for this hypothesis in which, for example, three versus fifteen target words per context were compared with CR effects on recall. Although context cues were helpful in aiding memory for the fifteen-words condition, the effect was even stronger for the three-words condition (Smith & Manzano, 2010), thereby providing evidence that cue overload (Watkins & Watkins, 1975) or fan effect (Anderson, 1974) does appear to exist.

Overshadowing hypothesis

In contrast to environmental context not being utilised as a retrieval cue during testing, the overshadowing hypothesis (see Matzel et al., 1985; Smith, 2013; Smith & Vela, 2001) relates to how the context is processed during learning. The overshadowing hypothesis is based on Glenberg's (1997) theory of environmental suppression and indicates that individuals generally automatically encode the environmental context unless they engage in additional conceptual processing. If the instructions given at the time of learning motivate participants to engage in greater conceptual processing, their limited attentional resources may lead to the suppression of the context so that the environment will be encoded to a lesser extent, or not at all. Therefore, the environment may only be weakly or not stored in memory and, as a result, contextual cues at test would not have a facilitating effect on memory.

At the same time, it could be hypothesised that an environmental change from learning to testing would also have a decreased detrimental effect on memory. The overshadowing hypothesis is also closely linked to research on cognitive/perceptual load and selective attention (Lavie, Hirst, De Fockert, & Viding, 2004). As the brain has only limited available cognitive resources, attention is selectively guided towards central, task-relevant information whilst ignoring task-irrelevant information. Lavie et al. (2004) make a distinction between perceptual and cognitive load and show that selective attention occurs at different times (early vs. late) depending on type of load. When perceptual load is high, people select early and distractor interference is reduced. In contrast, when cognitive load (e.g., working memory tasks) is high, people select late and distractor interference is high.

Research postulating that the environment can be distracting, has demonstrated, for example, that individuals automatically close their eyes during a difficult task (Doherty-Sneddon, Bruce, Bonner, Longbotham, & Doyle, 2002; Doherty-Sneddon & Phelps, 2005; Glenberg, Schroeder, & Robertson, 1998). Also, during memory recall, performance can be improved by either looking away from the environment (Glenberg et al., 1998) or even closing eyes completely (Perfect et al., 2008). These studies provide support for the overshadowing hypothesis in that participants are less likely to fully process the environment if presented with a cognitively demanding task, and would in turn not benefit from CR.

The above-outlined theories are often referred to within the CR literature and provide a plausible explanation as to why the extent to which context is used as a retrieval cue can differ. The CR research outlined in this review so far have exclusively used a *recall* procedure as a measure of memory. In contrast, studies that have included a test of *recognition* have also provided inconsistent findings, and this contrasting cognitive mechanism is discussed in the following section.

Recognition memory

Research including recognition tests generally investigate recognition memory on a Yes/No or Old/New test and often adopt the ICE (item, context and ensemble information) theory (see Murnane, Phelps, & Malmberg, 1999). Similar to the encoding specificity principle (Tulving & Thomson, 1973), ICE theory suggests that information encoded during learning consists of both item and contextual information, but also information related to the ensemble of the two. During such recognition tests, encoded contexts would then be presented again either with a new or old stimulus. This may lead to a 'context-dependent recognition'. When the encoded context is shown again during testing, this should lead to a feeling of familiarity. This feeling of familiarity may now be wrongly attributed to the stimulus, whether it is old or new. In other words, it is likely that there is an increase of correct identification of the old stimuli, whether those are shown with the original or a new context; however, false alarms also increase. The remember-know paradigm is also somewhat relevant here as it differentiates between individuals consciously remembering, and simply knowing that something is familiar (e.g., Mantyla, 1997; Tulving, 1985; Wixted & Mickes, 2010). In contrast, a 'context-dependent discrimination' occurs when old stimuli are shown with their original context—that is, the item-and-context-ensemble—which would increase correct identification of old (target) items but not false alarms to foils.

Empirical evidence provides some support for context-dependent recognition of words and items, with an increase in false alarms to foils when these are shown in studied contexts (Hockley, 2008; Hockley, Bancroft, & Bryant, 2012). Context-dependent discrimination, however, is supported by some studies (Macken, 2002; Tibon, Vakil, Goldstein, & Levy, 2012) but not others (Hockley, 2008). In fact, the inconsistency of this effect is even more-strongly shown in Hanczakowski, Zawadzka, and Coote (2014) where, over the course of three experiments, context-dependent discrimination occurred in two experiments but not a third. Hockley's (2008) findings could explain these mixed results: context-dependent discrimination seems to be mediated by whether participants have been explicitly asked to create an associative connection between study items and context, an effect that is called interactive encoding during testing.

More-general findings of studies investigating context effects in recognition are similarly inconclusive. Whilst some have shown a physical and mental CR benefit (Geiselman & Bjork, 1980; Phelps, 1998; Smith, 1985), though at cost of increased false recognition (Yamada, Nabeta, Oka, & Chujo, 2009), others have failed to find any significant difference between the CR and control condition (Godden & Baddeley, 1980; Hockley, 2008; Jacoby, 1983; see also Smith & Vela, 2001, for a review). Smith, Glenberg, and Bjork (1978), who included both a recall and a recognition procedure, found a CR benefit for the former but not the latter condition. That said, in Smith and Vela's (2001) meta-analysis, there was no significant difference between recognition and recall, indicating that when all studies were taken together the effect size of recall (d = .29) was actually very similar to the effect size of recognition (d = .27). At first glance this seems surprising; however, when looking closer at the experimenters' inclusion criteria, it is perhaps apparent as to why this is: studies were included that investigated recognition not only of words but also of faces.

Conceivably, the reason for this null result is that these two types of recognition cannot be adequately combined, as face perception is arguably very different to that of words, and different brain regions are involved for these different processes (e.g., see Davies-Thompson, Johnston, Tashakkor, Pancaroglu, & Barton, 2016). Unknown faces only seen once are still difficult to recognise later (e.g., Bruce, Henderson, Newman, & Burton, 2001; Davis & Valentine, 2009; Megreya & Bindemann, 2009), especially as the face is usually not seen in exactly the same way again, but perhaps in a different pose, lighting and with a different facial expression (e.g., Davies & Milne, 1982). Although this would be comparable to words being shown in a different font type (e.g., Naveh-Benjamin & Craik, 1995), Smith and Vela (2001) did not make this distinction. Further, whilst faces in CR studies are typically unfamiliar to participants, words tend to be familiar. There are relatively few experiments that have included unfamiliar verbal material (non-words), and those which have provide evidence that context changes only impact on word recognition for familiar, not unfamiliar words (e.g., Russo, Ward, Geurts, & Scheres, 1999). This reasoning is consistent with research indicating that, generally, more-difficult tasks tend to benefit from context cues to a greater extent than easier tasks (e.g., Cutler, Penrod, & Martens, 1987). Also, the familiarity of words allow those to be stored in memory independent of contextual information (Dalton, 1993), thereby reducing the necessity of context cues. Therefore, whilst CR effects for face recognition tend to occur (e.g., Evans, Marcon, & Meissner, 2009; Rainis, 2001; Shapiro & Penrod, 1986), this effect is less likely for word recognition (e.g., Godden & Baddeley, 1980; Smith et al., 1978).

Two theories can help to further explain why word recognition may not benefit from CR. Firstly, the continuity hypothesis (Tulving & Watkins, 1973) indicates that although retrieval processes are the same in recall and recognition tests, free recall requires more intensive involvement of these processes. Whilst retrieval cues—in this case, contextual cues—are effective in facilitating memory during free recall, effectiveness is reduced in recognition tests as the to-be-remembered word is already present. Secondly, the outshining hypothesis (Smith, 1988, 1994; Smith & Vela, 1986), as discussed earlier in this review, also provides a plausible explanation. As the to-be-remembered item is already present during the recognition test, the need for using contextual cues to aid memory is eliminated.

Whilst research concerning CR effects on recall and recognition of verbal memory is theoretically interesting, and could imply that findings may be applicable to a real world setting, the methodologies tend to lack ecological validity and, from those alone, it would be unclear as to whether CR effects would actually occur in more-realistic settings. Some studies with greater ecological validity have investigated CR effects on eyewitness memory, and as these are more relevant to the current thesis, are described in the following section.

Eyewitness memory

Exploring CR effects on eyewitness memory is of obvious forensic value, and findings are easily applicable during a crime investigation. The methodology of studies that have investigated this mechanism generally make use of a video clip of a staged crime scene (e.g., Wong & Read, 2011), or a staged event (e.g., Dando, 2013), and incorporate an interval of two days (Wagstaff, Wheatcroft, Caddick, Kirby, & Lamont, 2011) up to a week (Wong & Read, 2011) before participant-witnesses are asked to recall information about the presented 'crime'. To explore context effects, participants either physically return to the environment in which they have previously been tested (Hammond, Wagstaff, & Cole, 2006; Wong & Read, 2011), or they are interviewed via mental context reinstatement (MCR), a now well-known technique with which the environment is mentally visualised and then recalled (Dando, Wilcock, & Milne, 2009; Wagstaff, Cole, Wheatcroft, Marshall, & Barsby, 2007).

Findings in this research area generally show an increase in correct recall following context cues (Dando, Wilcock, & Milne, 2009; Hammond et al., 2006; Wagstaff et al., 2007; Wong & Read, 2011), with little or no increase in false recall (Davis, McMahon, & Greenwood, 2005; Emmett, Clifford, & Gwyer, 2003; Memon & Bruce, 1995; Milne & Bull, 2002). Even recall of a license plate can be significantly increased via MCR interview (Emmett, Clifford, Young, Kandova, & Potton, 2006). It is perhaps worth mentioning that the MCR instructions tend to work only for individuals without Autism Spectrum Disorder (ASD), and those with ASD only benefit from a physical CR (Maras & Bowler, 2012). This is thought to be due to individuals with ASD having difficulty in engaging in mental 'time travelling' (Lind & Bowler, 2010), an activity that seems to be required during memory recall (Tulving, 1985) than recognition. To summarise, research generally suggests that in a real crime investigation, eyewitness memory of the event should be improved by reinstating contextual information. Individual differences of the eyewitness, such as ASD, would infer whether context should be reinstated physically or mentally.

Outside of the laboratory, there are already CR techniques available to facilitate and improve eyewitness memory recall. Witnesses are commonly interviewed via the Cognitive Interview (CI; e.g., Geiselman, Fisher, MacKinnon, & Holland, 1985; Wells, Memon, & Penrod, 2007), an interviewing technique for witnesses of crimes aimed at promoting their memory recall. The CI consists of several mnemonics, one of which is MCR. As part of an MCR interview, witnesses are given similar instructions as within the academic literature: they are asked to mentally reinstate the environmental and psychological context, along with other physical states (e.g., smells, sounds). The literature repeatedly shows that MCR elicits a greater recall of correct information compared to a standard ('question and answer' type) police interview, with little or no increase in incorrect information (Memon, Meissner, & Fraser, 2010). The effectiveness of MCR has been demonstrated in both field and laboratory studies (using vignettes, videos, staged events; e.g., Cutler et al., 1987; Dando, Wilcock, & Milne, 2008; Hammond et al., 2006). A further benefit is that jury members' impressions on the suspects' guilt are generally not influenced by use of the CI and MCR as they would be, for example, with a hypnosis-type technique (Kebbell, Preece, & Wagstaff, 1995). As well as the 'report everything' mnemonic, the MCR is seen as the most valuable mnemonic to facilitate eyewitness recall (Davis et al., 2005; Milne & Bull, 2002).

Despite these positive findings, there are also some practical issues related to the CI and the use of the MCR technique. Repeatedly, it has been reported that police officers do not actually use the CI to interview witnesses; there is evidence that this is the case in the UK (Clarke & Milne, 2001; Clifford & George, 1996; Dando Wilcock, & Milne, 2009), in the US (Fisher, Geiselman, & Raymond, 1987; Schreiber Compo, Hyman Gregory, & Fisher, 2012) and in Canada (Wright & Alison, 2004). When more-closely inspecting the mean number of times in which a sample of police officers actually used the MCR technique per interview (N = 26), it appears to be only 8% of the time (Schreiber Combo et al., 2012). One of the issues is that the full CI technique, and the MCR in particular (Clarke & Milne, 2001), are considered to be time consuming (Kebbell & Wagstaff, 1999). A further issue is that it is often used incorrectly (Clarke & Milne, 2001; Dando et al., 2008), an outcome that is likely to have a detrimental effect on memory recall (Rosenbluth-Mor, 2001, cited in Dando et al., 2008). So, although the MCR has received much supporting evidence highlighting its beneficial effect, police officers seem either unwilling or unable to make effective use of the technique.

To avoid or at least limit these issues, Sketch MCR can alternatively be used. With this technique, witnesses are asked to draw a rough sketch of the environmental context in which the event/crime took place. This technique not only produces correct recall rates similar to the full MCR interview, but both of these techniques are superior to not reinstating the context at all (Dando, Wilcock, & Milne, 2009) for both younger (19-39 years; Dando, Wilcock, & Milne, 2009; 18-45 years; Dando, Wilcock, Milne, & Henry, 2009) and older adults (> 65 years; Dando, 2013). A further promising result needs to be highlighted: there is a decreased number of confabulations with Sketch MCR (e.g., Dando, Wilcock, & Milne, 2009; Dando, Wilcock, Milne, & Henry, 2009). Confabulations are defined as erroneous memories where individuals unintentionally fabricate or misinterpret events. During a crime investigation, it is clearly beneficial to ensure that witnesses do not create confabulations, highlighting an advantage of using Sketch MCR. However, the issue regarding time constraints may remain as the mean duration of interview was only reduced by 17% when using Sketch MCR relative to MCR (Dando, Wilcock, & Milne, 2009).

Focused meditation (FM)—with which participants listen and follow audiotape instructions of focused breathing exercises—can also have positive effects on memory recall (Wagstaff, Brunas-Wagstaff, Cole, & Wheatcroft, 2004; Wagstaff, Brunas-Wagstaff, Knapton, et al., 2004). Although this technique would increase the required interviewing time further (by approx. 1.5 minutes), combining FM with MCR is even more successful in increasing eyewitness memory recall than MCR alone (Wagstaff et al., 2007, 2011). This technique also neither increases errors nor inflates confidence (Wagstaff et al., 2007). Further, it is also not only effective with adults but also with children, again with no increase in erroneous recall: both the MCR and FM technique have been successfully tested with 6- to 7- (Wagstaff et al., 2011) and 11- to 12-year olds (Hammond et al., 2006). Unfortunately, in these experiments with child witnesses, techniques were only tested separately and not in combination. Therefore, it seems crucial to investigate the techniques in combination, to explore whether there would also be an additive effect for children as there has been for adults (e.g., Wagstaff et al., 2007, 2011). It perhaps needs to be highlighted here that whilst the MCR technique led to greater recall than the FM technique, it is also associated with increased levels of false confidence (e.g., Hammond et al., 2006). This would be a problem in a real crime situation as witness confidence can have an influential impact on jurors' perceptions of witness credibility (Wells, Ferguson, & Lindsay, 1981; Wells, Lindsay, & Ferguson, 1979). Therefore, it seems beneficial to combine MCR and FM to not only increase recall further but to also avoid inflated confidence as previously demonstrated by Wagstaff et al. (2007).

In addition to the aforementioned studies (esp. Hammond et al., 2006; Wagstaff et al., 2011), there have been many experimental studies that have also investigated the MCR technique with children (e.g., Dietze, Powell, & Thomson, 2010; Dietze, Sharman, Powell, & Thomson, 2013; Drohan-Jennings, Roberts, & Powell, 2010; Hershkowitz, Orbach, Lamb, Sternberg, & Horowitz, 2001; Shrimpton, Oates, & Hayes, 1998). As children tend to have worse recall due to retrieval deficits (Ackerman, 1985), and children's rate of forgetting tends to be greater than that of adults (Brainerd, Reyna, Howe, Kingma, & Guttentag, 1990; Flin, Boon, Knox, & Bull, 1992), theoretically it seems that children would benefit from context cues to a greater extent than adults. Whilst some studies indicated that MCR instructions are not effective with children (Dietze et al., 2013), or only effective in subsequent cued but not free recall (Dietze et al., 2010), the majority suggest that it can be a successful technique for children (e.g., Bowen & Howie, 2002; Dietze, Powell, & Thomson, 2012; Dietze & Thomson, 1993; Drohan-Jennings et al., 2010; Hayes & Delamothe, 1997; Hershkowitz et al., 2001; Memon, Holley, Wark, Bull, & Köhnken, 1996). The physical CR on its own does not seem to facilitate children's recall (e.g. Shrimpton et al., 1998); however, providing photo retrieval aids in combination with MCR instructions (Aschermann, Dannenberg, & Schulz, 1998) or MCR instructions on their own (e.g., McCauley & Fisher, 1995) can increase correct recall as well as reduce erroneous information.

Although there has been some concern that MCR may only work with older but not younger children (below 7 years; e.g., Dietze & Thomson, 1993; Hayes & Delamothe, 1997; Naka, 2012), other laboratory studies indicate that it can be successful for a variety of age ranges: 4 to 6-year-olds (Bowen & Howie, 2002), 5 to 11-year-olds (Dietze & Thomson, 1993; Hayes & Delamothe, 1997; Memon et al., 1996), 5 to 7- and 9 to 11-year-olds (Dietze & Thomson, 1993; Hayes & Delamothe, 1997; Memon et al., 1996), 6 to 7-year-olds (Drohan-Jennings et al., 2010), 6- and 11-year-olds (Dietze & Thomson, 1993), 6-year-olds, 8-year-olds and 11-year-olds (Dietze et al., 2012). Field studies have also found that 4 to 6- and 7 to 13-year-olds provide more information following the MCR technique during free recall compared to a standard structured interview (Hershkowitz et al., 2001; Hershkowitz, Orbach, Lamb, Sternberg, & Horowitz, 2002).

A further aspect of eyewitness memory is facial memory, or more-specifically identification of perpetrators and suspects from their face. As CR has been used to improve both verbal and eyewitness memory, research has also investigated whether contextual cues can be used to improve face-recognition rates. The relevant literature is outlined in more detail below.

Facial memory

As with verbal memory, context to improve facial memory can be defined in various ways. The literature indicates that this can be the physical environment in which a face had been encoded (Davies & Milne, 1985; Wagstaff, 1982), the environmental background of a facial photograph (Beales & Parkin, 1984; Davies & Milne, 1982; Gruppuso, Lindsay, & Masson, 2007; Hanczakowski, Zawadzaka, & Macken, 2015; Memon & Bruce, 1983; Parkin & Haywad, 1983; Reder et al., 2013; Thomson, 1981; Thomson, Robertson, & Vogt, 1982), the accompanying descriptive phrases next to a target face (Bower & Karling, 1974; Kerr & Winograd, 1982), clothing (Brutsche, Cissé, Delegeise, Finet, Sonnet, & Tiberghein, 1981; Sporer, 1993; Thomson, 1981; Thomson et al., 1982), and the accompanying face next to a target (Watkins, Ho & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977). Although face recognition has been facilitated through context reinstatement effects in the aforementioned cases, the extent of improved face identification varies. This is possibly due to the fact that many different factors have been defined as *context* and that these contexts may not actually be adequately comparable.

It is important to point out that CR is generally only useful in facilitating recognition of novel rather than familiar faces (Dalton, 1993; Davies & Milne, 1982; Russo et al., 1999). This is possibly due to the fact that faces of different familiarity are processed in different memory stores: whilst familiar faces will be stored in long-term memory, novel faces tend to be represented in episodic memory (Bruce, 1979, 1982; Ellis, Shepherd, & Davies, 1979). Familiar faces are also highly recognisable regardless of visual angle, lighting or other changes (see Johnston & Edmonds, 2009 for a review), even when a photograph (Bruce & Young, 1998) or CCTV footage (Burton, Wilson, Cowan, & Bruce, 1999) of a familiar face is of poor quality. Therefore, it is unsurprising that context cues are not needed to aid recognition of a familiar face.

The majority of the experiments on CR effects on face recognition mentioned above can be argued to have limited ecological validity (e.g., descriptive phrases next to a target face), and are therefore perhaps not adequately generalisable to a real life situation (Brunswik, 1956). Those studies with greater practical importance have investigated context effects concerning eyewitness identification, where participants are asked to identify the target ('perpetrator') amongst a choice of alternatives. Eyewitness identification can be poor (Brigham, Maass, Snyder, & Spaulding, 1982; Loftus, 1979; Penrod, Loftus, & Winkler, 1982; Yarmey, 1979), which can in turn lead to either false identification and/or the actual perpetrator not being arrested. This outcome highlights the need for identifying techniques of how to improve and facilitate eyewitness memory, such as with the use of CR techniques.

For example, Thomson et al. (1982) showed images of a target committing a crime and manipulated the background context as well as the target's identity. As expected, target faces were more recognisable when shown in the same as compared to a different background. However, as argued by Bruce (1982), a problem with this kind of experiment is that it is difficult to distinguish whether face recognition ability is actually involved, or whether what is being measured is simply 'picture recognition'. Picture recognition refers to discrimination between novel and previously-presented images: that is recognition of the same object presented in the same way

—be it a house, hat, mountain or face. In contrast, face recognition refers to the capability of discriminating novel from previously-presented faces, especially when what is seen has changed in some way between study and test (e.g., changes in facial expression, head pose and lighting). Also, as properties of a visual scene such as facial expression and pose reliably facilitate face recognition (Ellis & Deregowski, 1981; Walker-Smith, 1980), it is crucial to replicate Thomson et al.'s (1982) findings with stimuli that present the previously-presented faces with a different expression and pose. Indeed, Davies and Milne (1982) replicated these findings, thereby implying that so-called 'structural coding' is occurring and that CR techniques can reliably assist face recognition.

It is perhaps worth mentioning that the above experiments (e.g., Davies & Milne, 1982; Thomson et al., 1982) seem to be quite far removed from a real life situation, as static pictures of a perpetrator committing a crime cannot have the same emotional impact. To improve ecological validity further, line-up (or identity parade) studies are useful to explore; although still lacking the same emotional impact as a real crime, these tend to be a little closer to real life. In line-up CR studies, participants are better-able to correctly identify the target when tested in the same rather than a different room where the initial face encoding occurred (Evans et al., 2009; Wagstaff, 1982; Wong & Read, 2011). Similar to the research outlined in previous sections, MCR carried out prior to line-up task is also effective in increasing recognition 'hits' (Cutler et al., 1987; Malpass & Devine, 1981), thus rendering physical reinstatement of the context unnecessary.

Even-more realistic studies have incorporated actors: for example, in Krafka and Penrod's (1985) experiment, a target individual entered a store in which he paid with a traveler's cheque, talked to the sales assistant and exited. Following this, the sales assistant was asked to identify the target from a line-up and was interviewed via MCR technique before the line-up or not. A marginally significant effect of MCR was found. This indicates that although still effective, MCR may not be as effective in real life situations as it is in the laboratory. Similarly, in Smith and Vela's (1992) staged experiment, only physical CR was found to be effective, not MCR. A reason for this finding may be the fact that participants were familiar to the environment in which the staged event occurred (this was their usual classroom). A familiar environment is likely to automatically activate memories related to events other than the experiment, and so participants are less likely to benefit from the environment as a retrieval cue (see Hupbach, Gomez, & Nadel, 2011). This result would also appear to be in line with the cue-overload principle (Watkins & Watkins, 1975), one which indicates that the likelihood of recalling the target memory declines with increasing number of competing targets in memory.

It should perhaps be noted that some of the experiments have provided details about the event as part of the MCR interview (e.g., Malpass & Devine, 1981), to give rise to an interviewing technique which is suggestive in nature (i.e., not coming from the memory of the eyewitness) and could have influenced results either by increasing memory recall due to the provision of details as opposed to the witness's own memory, or by potentially inducing confabulations. In other experiments, physical evidence from the event is additionally shown (e.g., Clifford & Gwyer, 1999; Krafka & Penrod, 1985), and therefore goes beyond the sole memory of the witness. This methodology provides a potential issue with real world applications, as in a criminal investigation it may not be feasible or appropriate either to provide physical cues or to employ suggestive interviewing techniques.

CR effects have also been considered as part of a meta-analysis of eyewitness and more-general facial identification studies which aimed to identify variables that facilitate recognition (Shapiro & Penrod, 1986). Aside from CR, the meta-analysis included a variety of other factors such as facial distinctiveness as well as sex, age and race of the target, depth of processing strategies, retention interval, facial recognition training and many more potentially-relevant variables. Analysis indicated that CR had the largest beneficial effect on correct facial identification; false identification also increased, albeit to a lesser extent. Although this finding stands in contrast to two experiments that did not show an increase in false identification (Krafka & Penrod, 1985; Smith & Vela, 1992), it mirrors previous research where greater false identification occurred not only in facial memory (Malpass & Devine, 1981; Rainis, 2001; Searcy, Bartlett, Memon, & Swanson, 2001) but also verbal memory (e.g., Hockley, 2008; Hockley et al., 2012), and this could cause concern in a real crime situation. The increased false identification rates may be due to the fact that participants may sense an increased 'feeling of knowing' (Schacter, 1983), or familiarity, as opposed to actual recollection, which in turn may be wrongly associated with recognising the target, thus resulting in a higher number of false alarms (Mantyla, 1997; Tulving, 1985; Wixted & Mickes, 2010).

Additionally, it is also worth mentioning that Shapiro and Penrod's (1986) meta-analysis revealed that the CR effect was smaller in field compared to laboratory studies. This indicates that in a real-world setting, CR may not be as effective as in the laboratory.

Rather than using physical environment as context, or asking eyewitness-participants to mentally reinstate the environmental context (e.g., Davies & Milne, 1985; Rainis, 2001; Wagstaff, 1982; Wong & Read, 2011), physical characteristics of the actual target identity could also act as context cues. In Cutler et al.'s (1986) study, participants were shown a video of a robbery and subsequently attempted identification in line-up slides. Prior to the slides presenting the line-up, participants in the CR conditions were interviewed via the MCR technique, re-read their own description of the robber, and answered a series of questions and carried out a checklist regarding the physical appearance of the robber—such as hair and eye colour, hair style and body build. These context cues combined in this way were effective in significantly improving identification, but had only a scarce impact when assessed separately. This finding is in line with previous research that facilitated identification through either combining contextual cues (such as voice samples, posture and gait of the target, Cutler et al., 1987), or combining a physical and a MCR technique along with physical cues (Clifford & Gwyer, 1999).

Whilst the technique of re-reading one's own description of a target seemed to be successful in increasing correct line-up identifications when combined with other cues in Cutler et al.'s (1986) study, it was not effective on its own in Sporer's (2007) experiment. Although identification accuracy increased by 16% in Sporer (2007), this difference was not significant. The lack of significance here, however, could be partly due to an issue of low statistical power. Also, the result could have been affected by the lack of control over accurate versus inaccurate descriptions: if descriptions were inaccurate, the act of re-reading could have reinforced participants' erroneous memory and induced a feeling of familiarity; this process, in turn, could have been misinterpreted as familiarity of the incorrect target (Read, 1995). In fact, confidence ratings decreased in participants who re-read their own descriptions in Sporer's (2007) experiment. That said, in a real-life situation, accuracy of descriptions could also not be controlled for and so this argument may be redundant. Sporer's findings are also relevant to verbal facilitation versus verbal overshadowing paradigms: verbal facilitation (e.g., Itoh, 2005) implies that a verbal description about a stimulus (e.g., a face) *promotes* its subsequent recognition, whilst verbal overshadowing (e.g., Schooler & Engstler-Schooler, 1990) implies that it *hinders* recognition. In Sporer's case, it would seem that findings may not only be due to the mechanism (or reason) discussed above; they may also be driven by verbal overshadowing.

There has been a further notable experiment that has failed to provide support for MCR techniques for improving facial recognition. In Searcy and colleagues' (2001) study, participants engaged with an experimenter who then acted as a future target identity and were asked to identify him from a line-up one month later. The MCR technique was not successful in increasing target accuracy compared to a standard interview. It is unclear as to why the effect could not be reliably replicated. In this experiment, participants were not aware that they would later be asked to identify the experimenter that they had seen in the first session, which may explain the results. This scenario is different to a real-life crime as witnesses would be aware that they may be asked questions about the perpetrator and so are likely to engage in conscious (intentional) face encoding (Frowd et al., 2015). However, this would have unlikely been the case with the experimenter. Nevertheless, these results inform cases in which witnesses are unaware of the perpetrator's criminal intentions at viewing, and as a result an MCR interview may not be effective in increasing identification.

Factors mediating CR effects on facial memory

As witnessing a crime can in some cases be stressful and traumatising, it seems crucial that research into CR techniques also examines how emotions may mediate CR effects on face identification. Rainis (2001), for example, presented faces with distinctive backgrounds that were aimed to elicit certain emotions. A negative emotion was caused by background images such as concentration camps or car accidents, whilst positive emotions were evoked by images showing Christmas parties or paradise islands. Neutral backgrounds, such as a car park or shop window, were also included, designed to evoke no emotion at all. These photographs were shown again during face recognition, thereby acting as a physical CR. Findings indicated that there was a general benefit for CR techniques to improve face recognition. With regard to the influence of emotions, faces associated with contexts that elicited negative emotions led to lower correct identification rates and higher false identifications compared to both neutral and positive emotions. CR techniques, however, helped to increase hit rates; this was not to the same level as the other two conditions but greater than without CR techniques. In fact, whilst the same context was effective in this respect, using a semantically-similar context was even more effective in increasing recognition hits than the same context. As mentioned above, negative emotions also increased false alarms, and using the CR technique with the same context increased false alarms further. Semantically-similar contexts, however, helped to reduce those again.

To summarise and to apply these results to a real world setting, when negative emotions are elicited in witnesses, their ability to correctly identify the perpetrator is decreased and they are more likely to incorrectly choose an innocent person in a line up. Using a physical CR technique would be beneficial in increasing correct identification, but using a semantically-similar context instead of the actual context should be even more-effective. This outcome may be comparable to the face versus picture argument: using the same context may be similar to picture recognition, but using a different, but semantically-similar context, promotes a more conceptually-driven recognition (Bruce, 1982).

A further study, however, would appear to contradict these findings: in Brown (2003), the CR technique was the same as in Rainis' (2001) study—that is, background images were shown again during testing as a physical CR. This time, correct identification rates increased for neutral and unusual events only, without impacting on highly arousing conditions. The author explains these results in terms of stress leading to a distinction in how central and peripheral details are remembered: stress can narrow attention to the central details and thereby reduce attention to the peripheral details, that is, those in the environment (e.g., Easterbrook, 1959; Eysenck, 1976). If the environment is not suitably encoded, contextual cues can therefore also not be helpful in facilitating memory recall.

There may again be a general issue with ecological validity for both of the above studies (Brown, 2003; Rainis, 2001), as images presented on screens are unlikely to evoke the same kind or level of emotions as when witnessing a crime. Ready, Bothwell, and Brigham (1997) attempted a more realistic approach: participants were told that they would have to give a public speech and be subsequently critiqued by a psychologist. Participants viewed an example video in which this particular psychologist critiqued someone else's speech and did so in a harsh and unfair manner, thereby aiming to elevate participants' stress levels. CR techniques used to facilitate facial memory for the speakers and psychologists seen in the video were, however, not successful. This could be explained, as indicated in the earlier paragraph, in that stress narrowed participants' attention to the central details at a cost to the encoding of peripheral, environmental details. Unfortunately, no control condition was included in which no stress was evoked.

To the author's knowledge, no further research has investigated environmental CR effects on face identification as a function of emotion; a research niche that would seem to warrant further exploration, especially since findings are inconclusive. Also, one could argue that witnessing a crime would evoke negative emotions in the majority of cases, if not all, and therefore it would be more informative to also manipulate intensity of negative emotions.

Another aspect that has been shown to influence facial identification accuracy is disguise. Research in face perception clearly shows large influences of the appearance of hair (e.g., Ellis et al., 1979; Goldstein & Mackenberg, 1966; O'Donnell & Bruce, 2001; Walker-Smith, 1978) and the region around the eyes (e.g., Goldstein & Mackenberg, 1966; O'Donnell & Bruce, 2001; Walker-Smith, 1978), both during encoding and identifying a face. If witnesses are not able to encode these regions of the face because of the presence of a disguise, it could be hypothesised that a CR technique would be particularly effective in eliciting memories. This was investigated in an experiment which used videotapes of a robbery, with a subsequent line-up task (Cutler et al., 1987). To explore the effect of disguise, a hat was added to the target, a factor that reduced accuracy of target identification (Cutler et al., 1986, 1987; Shapiro & Penrod, 1986). As expected, the MCR was most effective when the face was highly disguised, but had a small effect when it was not. This finding is in line with the outshining hypothesis (Smith, 1988, 1994; Smith & Vela, 1986): the need for the MCR technique as a cue to recall is increased when the face is poorly encoded due to the disguise, whilst reduced when it is better encoded (no disguise).

Similarly, MCR is also more effective when the retention interval between viewing and testing is longer (2 weeks)—and memory is therefore worse—than following a shorter interval of time (2 days; Cutler et al., 1987). Finally, the MCR technique is also more effective in high-similarity line-up conditions (where the target individual is seen amongst many other members who closely resemble the target's appearance) compared to low-similarity conditions (where only a few faces resemble the target's appearance; Cutler et al., 1987), again indicating that this technique is more effective when the task is more difficult to complete.

With this in mind, one may also expect that CR should be particularly effective for faces from another race since they are generally not encoded to the same extent as those of one's own race; this effect is coined the own-race bias (ORB; e.g., Malpass & Kravitz, 1969; Meissner & Brigham, 2001). ORB may be due to the fact that faces of one's own race (or of a race with which we have considerable experience) are encoded more-qualitatively, with greater memory for facial characteristics and episodic or source information than cross-race faces (Meissner, Brigham, & Butz, 2005). Additionally, it is theorised that there is a lack of attention paid to individual features of faces from other races, especially those features that are important for recognition (Hills & Lewis, 2005, 2006). Therefore, it could be theorised here that a CR technique would be more effective for cross-race faces, given that the above-mentioned experiments found MCR to be especially effective when memory is worse (e.g., Cutler et al., 1987). However, the opposite has been shown to occur: discrimination accuracy of own-race faces in target-present and target-absent line-up tasks is facilitated through the MCR, whilst no such effect is evident for cross-race faces (Evans et al., 2009). Although at first glance surprising, this effect can be explained in different terms. As the MCR procedure in Evans et al. (2009) consisted of reinstating the semantic information rather than environmental context, it would seem that as semantic information (e.g., biographical information) is encoded ineffectively for cross-race faces, the binding of such semantic information with facial information may be reduced. The particular study that has explored this effect (Evans et al., 2009) was only able to include Hispanic participants who viewed Caucasian targets. This design makes it difficult to generalise to other races, although it seems likely that the effect would be the same, since findings on the own-race effect are similar across various different races (see Meissner & Brigham, 2001 for a meta-analytic review).

Considering that older adults tend to have worse memory for an event than younger adults (e.g., Balota, Dolan, & Duchek, 2000), the elderly should thereby, theoretically, benefit from contextual cues to a greater extent. Empirical evidence only partially supports this theory. In Wilcock, Bull, and Vrij (2007), both age groups did not benefit from CR techniques for a line-up with young-aged targets, and only older participants showed a CR effect for an line-up of old-age targets. In fact, line-up identification performance of older adults was increased to that of younger adults (Wilcock, Bull, & Vrij, 2007). In the research, this finding was limited to a target-absent line-up only (i.e., there was no effect for target-present arrays). Target-absent line-ups could be considered to be more challenging than target-present line-ups, and this would thereby be in line with previous findings concerning CR having stronger effects for more difficult tasks (e.g., Cutler et al., 1987). To the author's knowledge, only two further studies have investigated CR effects with older adults (Searcy et al., 2001; Memon, Hope, Bartlett, & Bull, 2002), and both of these papers have failed to find any benefit from CR—this was the case for both younger and older adults (discussed in more detail below). Therefore, it is crucial that more research investigates the use of CR techniques with older adults. Similarly, whilst CR effects with children have been investigated in eyewitness research more-generally (e.g., Dietze et al., 2010, 2013; Drohan-Jennings et al., 2010; as discussed earlier in this review), this has yet to be explored within research on face recognition.

Alternative uses of the MCR technique in facial memory research

The above research has focused on how reinstating the context, physically and mentally, can improve facial memory and identification. However, the technique could also be used to reduce occurrence of eyewitness memory errors. For example, eyewitnesses tend to be more likely to choose an innocent target from a line-up that they have previously seen in a mugshot album, a phenomenon called the 'mugshot exposure effect'. Therefore, Memon et al. (2002) aimed to use the MCR technique not to improve correct face recognition, but to reduce false choosing from a target-absent line-up following mugshot viewing. Yet, results revealed that false choosing was not affected by the MCR interview and was deemed unsuccessful by Memon et al. (2002). Considering past research, however, this finding is perhaps not surprising: although false recognition has not been affected by the MCR technique in two experiments (Krafka & Penrod, 1985; Smith & Vela, 1992), the majority of evidence suggests an increase in false identification (e.g., Rainis, 2001; Searcy et al., 2001; Shapiro & Penrod, 1986), as discussed in detail above. Therefore, the rationale for Memon et al.'s (2002) results seems predictable from past research.

The MCR interview has also been used to reduce the influential impact of misleading post-event information. Jenkins and Davies (1985) illustrated that an erroneous Photofit facial composite—that is, an image of a target face (a 'perpetrator' in a crime situation) that has been created from an eyewitness's memory—when shown between target viewing and a line-up, reduces recognition of the target. This could be explained with Loftus's (1979) alteration hypothesis which indicates that memory is permanently altered and replaced as a result from having been exposed to post-event, erroneous information. In contrast, the co-existence hypothesis (Bekerian & Bowers, 1983; Christiaansen & Ochalek, 1983) stipulates that both the original and the new, erroneous memory co-exist, but that the former is rendered inaccessible as a result. To eliminate this issue, the MCR technique can be effectively used prior to the line-up task (Bekerian & Bowers, 1983; Gibling & Davies, 1988). Whilst Loftus's (1979) alteration hypothesis is challenged by these results, the co-existence hypothesis (Bekerian & Bowers, 1983; Christiaansen & Ochalek, 1983) could indicate that the original memory is made accessible as a result of the MCR technique. This effect, however, has only been demonstrated with the archaic Photofit system, and it is therefore difficult to say whether it could also be replicated with newer, feature- or holistic-based composite systems (described below).

Aside from facilitating facial memory and reducing eyewitness memory errors, it is worth highlighting that with an MCR technique, memory for the context can also be improved. In Hanczakowski et al. (2015), contexts that had initially been reinstated to facilitate face recognition were later also found to be recognised significantly better compared to other contexts. This result could also be of practical value during a crime investigation if potentially important contextual information is more likely to be recalled as a result of the MCR technique.

Considering that positive context effects on face identification have been shown across the research literature, and the fact that MCR is already applied to facilitate eyewitness recall in practice, it is surprising that only a minor reinstatement of context would appear to be common practice prior to face composite constructions, and not a more-extensive one. This thesis looks at just such a notion, extensive recall of the context in an attempt to facilitate face construction. In the following section, a review of the facial composite literature is outlined, and it is then explained in greater detail how CR effects could potentially be applied to this forensic area.

Facial composites

Overview of facial composite systems

In a forensic context, facial composites are likenesses of a target face constructed from eyewitness memory. In a crime situation, it is often the case that, if no other evidence is available, witnesses and victims of crime are asked to construct a composite commonly 1-4 days after the crime (Frowd, Pitchford, et al., 2012). The resulting composite is subsequently published with the aim that the perpetrator can be identified by a member of the public or the police. In the UK, facial composites are not only used to provide intelligence for an investigation, but they can also be used as evidence (ACPO, 2009).

Prior to constructing a composite, eyewitnesses are interviewed via a shortened version of the Cognitive Interview (CI; see Frowd, 2011 for an in-depth review of the CI for facial composite construction). Best practice is that rapport is first built, to put witnesses at ease, a strategy which should also facilitate the subsequent memory recall (Collins, Lincoln, & Frank, 2002). Witnesses are then invited to mentally visualise the to-be-constructed face, and to freely recall as many details as possible, without guessing. Witnesses are not interrupted during this process as this could be detrimental to their memory recall. The practitioner usually takes written notes.

Following the CI, a number of techniques are available to construct composites (see Fodarella, Kuivaniemi-Smith, Gawrylowicz, & Frowd, 2015 for a more-detailed description). The earliest technique introduced is Artist's Sketch, which refers to a face being drawn by hand. More-specifically, following a witness's description of the face, a forensic artist creates an initial faint sketch which is then altered through the eyewitness's instructions. Commonly, the face would be drawn from the top of the head downwards, leaving individual features (e.g., eyes, nose, mouth) until the end. Forensic artists can also utilise reference materials, such as Samantha Steinberg's facial catalogue (2012), a reference source that contains images of facial features that may help in triggering the witness's memory (e.g., Kuivaniemi-Smith, Nash, Brodie, Mahoney, & Rynn, 2014). The sketch is mostly drawn with pencils or charcoal; however, colour can also be used. With developing technology, many artists prefer to use digital Sketch techniques, such as Adobe Photoshop, which enable easier amendments to be made to the face and quicker distributions of composites.

Sketch is still used in police forces today, and when compared to alternative systems, they are somewhat more identifiable than composites constructed with 'feature' systems but less identifiable than those from the holistic system EvoFIT (see below for a more-detailed description of holistic systems; Frowd et al., 2015). These findings can be explained by differences in procedures used across systems. EvoFIT is more-closely aligned to holistic face processing (e.g., Valentine, 1991;

Young, Hellawell, & Hay, 1987) than Sketch (discussed in more detail later in this review), which thereby leads to the best-quality composites. Using feature systems, constructors manipulate individual facial features in a sequential manner, whereas Sketch allows them to work on groups of features, making this process more in line with holistic face processing (Davies & Little, 1990; Laughery, Duval, & Wogalter, 1986).

As Sketch requires forensic practitioners to be highly artistically-skilled, 'mechanical' systems (e.g., Photofit, Identikit) were introduced to allow a wider range of practitioners to create facial composites. With these systems, eyewitnesses would be shown a range of individual facial features (hair, face shapes, eyes, etc.) printed on jigsaw-like pieces or transparencies from which they could select to build the target face. These features would be selected in isolation to each other (e.g., Shepherd & Ellis, 1996). However, poor correct naming rates are evident from composites constructed with manual and mechanical systems (Davies, van der Willik, & Morrison, 2000; Ellis, Davies, & Shepherd, 1978; Laughery, Duval, & Fowler, 1977; Laughery, & Fowler, 1980). Laughery et al. (1977) even found that Identikit composites constructed with the target 'in view' were just as poorly named as those constructed from memory—a result that would not be expected from theory (i.e., performance should be worse from memory, when the face was not visible during composite construction).

These negative findings are not surprising, however, when considering the way that we naturally remember and perceive faces (e.g., Valentine, 1991; Young et al., 1987) is different to the approach used by these systems when building a face. These systems have been developed based on the assumption that face recognition is a collection of individual analyses of facial features (Penry, 1971). However, faces tend to be perceived 'holistically'—that is, as a whole face—rather than through individual facial features (Richler, Cheung, & Gauthier, 2011; Tanaka & Farah, 1993; Valentine, 1991; Valentine & Endo, 1992; Young et al., 1987). Also, the configuration of individual features—that is, the physical distances between facial features—is important for face identification (e.g., Hole, 1994; Hole, George, Eaves, & Rasek, 2002; Lander, Bruce, & Hill, 2001; Tanaka & Sengco, 1997; see Cabeza & Kato, 2000 for a review), but the mechanical systems were limited as to the extent to which this aspect of the face could be altered. It has long been known that the low naming rates are a direct result of these limitations (see Frowd, 2017).

Mechanical systems have now been nearly entirely replaced by computer systems due to the aforementioned limitations (McQuiston-Surrett, Topp, & Malpass, 2006). Digitally constructing a face enables quicker amendments, such as re-sizing and re-positioning of features (Frowd, 2012), and quicker distribution. Also, a wider range of features have been made available in computer systems. Two main types of computer systems are available, feature and holistic. Feature systems (e.g., E-FIT, FACES, PRO-fit) are similar to the aforementioned 'mechanical' systems in that composites are built through selection of individual facial features. With the UK feature system PRO-fit, for example, practitioners use a witness's descriptions of the face and its individual features to select the appropriate characteristics within the system. Thereby, the number of features that are shown to witnesses is reduced from the full database (150-500 items per feature) to those items that fit the given descriptions (approx. 20 items). Once descriptions for each feature have been entered in the system, witnesses are shown an 'initial' composite face. In most cases, this face is a poor resemblance, and so the usual procedure is to improve the match by changing, re-sizing and re-positioning features. Each feature is worked on individually but within the context of the whole face, a technique that leads to more-accurate feature selection (e.g., Skelton, Frowd, & Speers, 2015; Tanaka & Farah, 1993), and this makes the technique qualitatively different to mechanical systems (i.e., that cannot conveniently show a range of features for selection in a whole-face context). PRO-fit composites are constructed in grey scale as colour does not seem to reliably affect identification[1] rates of resulting faces (Frowd, Bruce, Plenderleith, & Hancock, 2006).

Although composite likeness has improved from mechanical to computerised feature systems (Frowd et al., 2005), correct naming rates remain low at around 20% for a short retention interval (e.g. Davies et al., 2000; Frowd, Hancock, & Carson, 2004), and almost 0% following a more-ecologically valid interval of two days (Frowd et al., 2005, 2015; Frowd, McQuiston-Surrett, Anandaciva, Ireland, & Hancock, 2007). One of the reasons for these low recognition rates seems to be due to the ineffective construction of the face's internal features. Original work by Ellis, Shepherd, and Davies (1979) highlighted that photographs of familiar faces tend to be recognised better based on their *internal* (e.g., eyes, nose, mouth) rather than their *external* features (e.g., hair, ears). The eye region in particular seems to be important for identifying familiar faces (e.g., Goldstein & Mackenberg, 1966; O'-Donnell & Bruce, 2001). For unfamiliar faces, the opposite applies, that is, that external features (e.g., hair, face shape) are more helpful in revealing the person's identify than internal features. In contrast, composites from feature systems tend to be identified similarly by their internal and external features (Frowd, Skelton, Butt, Hassan, & Fields, 2011), thereby indicating that the internal region is not constructed more accurately than the external region for recognition (as is the case with facial photographs).

---

[1] The author acknowledges that the term 'identification' is different to 'naming'; however, this term is used within the field as an indicator of correct naming.

To overcome these issues, and thereby in an attempt to improve composite likeness and naming rates, the more-recent holistic systems (e.g., EFIT-V, EvoFIT, ID) take a more theory-driven approach with regard to how the brain perceives, remembers and recognises faces—that is, as a whole face as opposed to individual features (e.g., Valentine, 1991; Valentine & Endo, 1992; Young et al., 1987). For example, the EvoFIT system presents arrays of whole faces (or whole-face regions) and witnesses are asked to repeatedly select those that most resemble the target face. The characteristics of these selections are merged to 'breed' further faces from which to select. At the end of this process, a face is evolved, and individual facial features and holistic characteristics can be altered, such as texture, weight and personality (e.g., attractiveness, honesty, dominance). Another holistic system that is widely-used in the UK is EFIT-V (now EFIT-6). The procedure is somewhat similar to EvoFIT in that witnesses make repeated whole face selections from arrays of faces. If eyewitness memory is strong, individual facial features are shown within the whole face whilst other features are blurred to avoid distraction. If memory is found to be weak, arrays of faces are presented that differ with regard to shape and individual features from which the witness can select. Hair is added at the end of the procedure for witnesses with a good memory, or prior to other individual features for those with a poor memory. Other holistic characteristics (e.g., health, hostility, weight) of the face can subsequently be altered, similar to EvoFIT.

Therefore, whilst feature systems require detailed face recall, holistic systems enable composite construction mostly through face recognition, a task which is thought to be less challenging than the former (Davis, Sutherland, & Judd, 1961). In line with this approach, holistic systems also do not require a detailed verbal description of the face. This is an advantage as facial memory fades rapidly (Ellis, Shepherd, & Davies, 1980), and facial descriptions are generally inaccurate (Davies, Shepherd, & Ellis, 1978). In fact, these systems can even be used if the witness is unable to provide a face description at all (Frowd, 2012).

As a result, effectiveness of composites from holistic systems tends to be much greater than feature systems and Sketch (e.g., Davis, Sulley, Solomon, & Gibson, 2010; Frowd, 2010). Field studies are also in favour of holistic systems (cf. feature systems) revealing higher suspect identification rates (Frowd et al., 2010; Solomon, Gibson, & Maylin, 2012). The holistic system EvoFIT emerged in a recent meta-analysis (of exclusively laboratory studies) as the most effective system, with over four times greater identification rates (at 56%) compared to Sketch and feature systems (Frowd et al., 2015). Unfortunately, it was not possible to also include the EFIT-V system in this meta-analysis due to a lack of available data sets using a forensically-relevant paradigm. Using a procedure that is less forensically-relevant (short retention interval), Valentine et al. (2010) reveals that EFIT-V composites were correctly named at 20.3%. It is likely that these relatively low naming rates are due the procedure used in this particular experiment: participants created eight composites and were likely fatigued during this process. The main aim was not to investigate identification rates but morphing effectiveness (discussed in greater detail later in this review). The EFIT-V system has now changed greatly, to become EFIT-6, but to the author's knowledge no published research has been carried out with the new system.

To ensure rigorous research is undertaken in order to evaluate composite systems effectively, it is crucial to examine the methodology used. Best practice is proposed in the following section.

Methodology in facial composite research

Facial composite research should aim to mirror real life situations—a consideration which Frowd et al. (2005) coined the 'gold standard' procedure. In the laboratory, participants in the composite-construction stage are often referred to as 'participant-witnesses'. At this stage, it is important to ensure that participants are unfamiliar with the target, as witnesses in real life are much more likely to be unfamiliar with a perpetrator. There are different approaches as to how to present a target face, such as, showing a photograph of a face (e.g., Fodarella et al., 2017; Gawrylowicz, Gabbert, Carson, Lindsay, & Hancock, 2012; Hasel & Wells, 2007; Kehn, Renken, Gray, & Nunez, 2014), a video clip (e.g., Davis, Gibson, & Solomon, 2014; Davis, Thorniley, Gibson, & Solomon, 2016; Marsh et al., 2017; Valentine et al., 2010), or a staged event (e.g., Davies & Milne, 1985). For reasons of efficiency, most research utilises a photograph of a face. Participant-witnesses are shown the target face for a specified amount of time (often 60 seconds), and are asked to remember as many details about the face as possible. Although it could be argued that using video stimuli should lead to a more realistic approach, there is evidence of little difference in identifiability of composites constructed following a video and those following static images (Frowd et al., 2015). Thereby, the use of static images seems to make practical sense.

Face construction occurs immediately or after a specified delay following face encoding. Different delays have been used in the research literature: (a) no (or very short) delay (e.g., a few minutes; e.g., Davis et al., 2014, 2016; Frowd et al., 2004; Hasel & Wells, 2007), (b) a short delay of a few hours (e.g., Frowd et al., 2005; Gawrylowicz et al., 2012), (c) a delay of 24 hours (e.g., Kuivaniemi-Smith et al., 2014), (d) 48 hours (e.g., Frowd et al., 2005; Frowd & Fields, 2011), (f) or a week (e.g., Davies & Milne, 1985; Green & Geiselman, 1989). A very short or no delay leads to the best composite likenesses, presumably as memory is likely to still be strong (Frowd et al., 2015). However, as witnesses often construct composites 1-2 days or even longer after the crime occurred, longer delays should thereby reveal results with greater ecological validity.

To evaluate the identifiability of these constructed composites, there are a number of tasks that can be used (also see Frowd et al., 2005). These are, for example, (a) a sorting/matching task in which participants are asked to match composites to original photographs of the target face (e.g., Frowd, Nelson, et al., 2012; Gawrylowicz et al., 2012; Green & Geiselman, 1989; Kuivaniemi-Smith et al., 2014), (b) a ranking task (e.g., Green & Geiselman, 1989; Valentine et al., 2010), to indicate best and worst likenesses, (c) a constrained naming task, in which participants are presented with both composites and a list of target names from which to select (Frowd & Fields, 2011), (d) a likeness rating task, in which the composites are rated for likeness on a Likert-scale (e.g., Bruce, Ness, Hancock, Newman, & Rarity, 2002; Frowd, Bruce, Smith, & Hancock, 2008; Hasel & Wells, 2007; Koehn & Fisher, 1997), (e) a spontaneous naming task, in which participants are presented with composites from a specific set of identities (e.g., a particular soap)[2] and asked to name them (e.g., Davies et al., 2000; Frowd et al., 2011; Frowd & Fields, 2011; Fodarella, Brown, Lewis, & Frowd, 2015; Kovera, Penrod, Pappas, & Thill, 1997), and (f) a cued naming task, in which composites are shown for a second time for identification following the spontaneous naming task and presentation of the original tar-

[2] The author acknowledges that by specifying the target set to participants, the task is somewhat 'cued'. However, not providing such contextual information leads to very low naming rates (see Valentine et al., 2010), thereby making it difficult to detect any statistical differences between experimental conditions. Composites in the real world are generally accompanied by some contextual information (e.g., details of offender's build or accent), and therefore it could be argued that this is comparable to specifying general information about the set in advance.

get photographs (e.g., Frowd et al., 2008; Frowd & Fields, 2011). The technique which seems to mirror the real world scenario to the greatest extent, and thereby reveals more ecologically valid results, is spontaneous naming.

Techniques for improving composite quality

As composites are constructed from memory, they are prone to error. To overcome these inaccuracies, there are a number of ways in which composite quality—that is, recognisability—can be increased. One way is to improve the composite system, as has been the case greatly with E-FIT, Photofit and EvoFIT (e.g., Frowd, 2017). Other ways are to facilitate eyewitness memory prior to composite construction or to manipulate presentation of composites to improve correct naming. The following section outlines these other techniques.

One way of improving performance is at the interview stage, conducted immediately prior to face construction. Instead of solely using the standard CI prior to construction, a 'Holistic' CI (H-CI) tends to improve composite likeness and thereby increase its identifiability. With this technique, witnesses are asked to make whole-face judgments about the personality of the face (Frowd et al., 2008, 2013, 2015; Frowd, Nelson, et al., 2012). Following the CI, they are asked to reflect silently on the personality of the face for 60 seconds, before rating personality traits (e.g., intelligence, friendliness, kindness) on a three-point Likert scale. Following this additional mnemonic, a composite is constructed as normal.

Face recognition may be enhanced with this technique as attributing traits promotes holistic facial recall (e.g., Davies & Oldman, 1999), whilst the standard CI promotes face description through feature recall. Featural processing, however, can interfere with face recognition (e.g., Dodson, Johnson, & Schooler, 1997), a factor that is important to consider during composite construction (e.g., Frowd & Fields, 2011). The H-CI is likely to cause a shift in this processing strategy towards a holistic approach, which—in contrast to a featural approach—should aid recognition (Davies & Oldman, 1999). Also, the H-CI may lead to a generally more in-depth involvement with, and analysis of, the face than with the standard CI. As a result, internal and external features are of better likeness following the H-CI (Frowd et al., 2008). As internal features are an important driver of face recognition (e.g., Ellis et al., 1979), it is likely that this thereby improves the resulting composite likeness (Frowd, Bruce, McIntyre, & Hancock, 2007). Further, other aspects of the constructed face, age and distinctiveness, also become more accurate following the H-CI; both are factors that facilitate face recognition (e.g., Ellis, 1986) and thereby composite effectiveness (Frowd et al., 2008). The H-CI does not only improve composites from holistic but also feature systems (Frowd et al., 2015). Although surprising at first, this outcome seems to be due to the fact that feature systems promote *some* holistic processing: facial features are judged within the context of the whole face, and so face recognition is engaged to some extent. In fact, Skelton et al. (2015) demonstrates that the feature system PRO-fit benefits from presenting the facial context during individual feature selection, thereby producing more-recognisable composites. Further support derives from findings regarding the use of H-CI for Sketch composites. Sketch seems to be a more recall-based technique, as the artist is relying on the witness's featural recall to a greater extent (cf. E-FIT/PRO-fit), and as a result the H-CI technique is not effective for Sketch composites (Stops, unpublished; cited in Frowd et al., 2015).

There are various techniques to improve recognition of composites that have already been constructed (for a review, see Frowd, 2017). One way is to present these facial images as a caricature (e.g., Frowd, Bruce, Ross, McIntyre, & Hancock, 2007; Frowd, Skelton, Atherton, Pitchford, Bruce, et al., 2012). Relative to an average face, positive caricatures exaggerate distinctive facial features, the spacing between features and the overall face shape, whilst negative caricatures reduce these differences (for more details on 'face space' models in relation to caricatures see Lewis, 2004; Lewis & Johnston, 1998, 1999; Valentine, 1991, 2001). For composites, caricature has been applied to the shape (cf. texture) information in the face. The finding (described below) is that caricature can be used to improve correct naming of composites, which mirrors research using facial photographs (e.g., Benson & Perrett, 1991; Lee, Byatt, & Rhodes, 2000; Lee & Perrett, 1997) as well as line drawings of faces (e.g., Ellis, 1990; Rhodes, Brennan, & Carey, 1987).

The research on caricaturing composites by Frowd, Bruce, Ross, et al. (2007) initially found minimal improvement in correct naming when a face was exaggerated at one level of caricature, a null result that applied overall to Sketch, E-FIT, PRO-fit and EvoFIT composites. However, a large increase emerged when these representations were presented as a sequence. Whilst facial information was most conveniently presented for recognition as an animation, it is not the animation (i.e., the motion) that was found to drive the effect, as a static array of caricatures was found to be just as effective in facilitating recognition. It could be theorised that caricatures make composites look more distinctive (Lee et al., 2000), which may in turn increase identification by exaggerating prominent features. At the same time, the sequence developed by Frowd, Bruce, Ross, et al. (2007) also contained negative caricature, a representation where features were not exaggerated (as with positive caricature), but diminished. Thereby, a further reason for the increased correct naming rates may be that caricaturing reduces the appearance of inaccuracies in composites, error introduced as a result of constructing a face from memory. As mentioned already, such inaccuracies are related to individual facial features as well as the spacing between features, the presence of which impacts negatively on recognition of the face (e.g., Frowd et al., 2005, 2015). A sequence of caricatures may thus help to minimise the impact of these inaccuracies and thereby improve recognition (Frowd, Bruce, Ross, et al., 2007; Frowd, Skelton, Atherton, Pitchford, Bruce, et al., 2012). This observed anti-caricature effect for composite recognition seems to be somewhat different to that of real faces, where anti-caricatures are not well recognised (e.g., compared to lateral ones, Lewis & Johnston, 1998).

A further manipulation of completed composites is to either vertically stretch these images or to present them side-on, thereby stretching the composite face in a perceptual way (e.g., Davis, Simmons, Sulley, Solomon, & Gibson, 2015; Fodarella & Frowd, 2013; Frowd et al., 2014). Vertical stretch is defined here as an affine facial-image transformation whereby the height of the face is stretched linearly (i.e., with the width remaining unchanged). This transformation technique changes the configural information by enlarging or compressing distances between features (e.g., the distance between eyes and nose), as well as featural information by increasing the relative size of individual features. Whilst it does not impact on recognition of photographs of faces (e.g., Hole et al., 2002)—presumably as these stimuli are highly recognisable to start with—, composites tend to be significantly better identified when presented in this way (Davis et al., 2015; Frowd et al., 2014). Similar to caricatures, it can be theorised that altering composites in such a way may serve to minimise some of the facial errors: as mentioned above, stretching changes the featural and configural information of facial composites, and as a result, their inevitable erroneous facial information may appear to be less prominent. Frowd et al. (2014) showed that the beneficial effect is particularly strong for composites that had been constructed following an H-CI (described above). This finding would seem to suggest that as H-CI composites would be of better-quality with regard to configural information, stretching the face may, above all, reduce errors related to the individual shape and placement of facial features, facilitating recognition.

Concealing erroneous featural and configural information in feature and holistic composites has also been investigated in more detail in the research literature (Fodarella & Frowd, 2013; Frowd et al., 2014). Adding a low level of Gaussian blur to obscure details of individual features can increase identification of holistic EvoFIT composites, but reduce naming of feature PRO-fit composites (Frowd et al., 2014). This finding could be due to the theoretical assumption that feature systems contain more-accurate featural than configural information, whilst the opposite would be the case for composites from holistic systems, thus allowing feature errors to be less apparent. With regard to manipulations likely to conceal configural information, horizontally-misaligning composite faces (about the vertical mid nose level) can increase naming of poorly-constructed (but not well-constructed) EvoFIT composites (McIntyre, Hancock, Frowd, & Langton, 2016). Similarly, poor composites can also be made more-identifiable when sunglasses are added (Brown et al., 2018; Fodarella & Frowd, 2013; McIntyre, Frowd, Bruce, & Hancock, 2010). This finding is unsurprising since the eye region is important for familiar-face recognition (Goldstein & Mackenberg, 1966; O'Donnell & Bruce, 2001), and thereby, if this region is constructed inaccurately, composites are unlikely to be recognised well. Therefore, concealing this relatively inaccurate eye region reliably aids the recognisability of composites, allowing the face to be recognised using other available facial cues (Brown et al., 2018; Fodarella & Frowd, 2013; McIntyre et al., 2010). Along with concealing the eyes with sunglasses, hats can also be added to conceal erroneous hair, again to improve composite naming (Brown et al., 2018).

This latter technique is of potential practical benefit for cases in which the perpetrator has altered their hair following the crime.

So far, all the techniques outlined in this section are useful for individual composites. However, a further manipulation has been developed when more than one composite of the same identity has been constructed, such as may be the outcome if multiple witnesses construct a composite of the same face. Due to individual differences, some individuals are able to produce more-accurate composites than others. To overcome the issue of not knowing who would create the best composite, it is advisable to ask witnesses to construct a composite independently of each other, and to subsequently 'morph' these multiple composites into a single, averaged composite face (Bruce et al., 2002; Davis et al., 2010, 2015; Hasel & Wells, 2007; Valentine et al., 2010). The resulting morphed composite tends to be recognised as well as the most-identifiable composite in the set, or even better (Bruce et al., 2002). This is the case for composites created with feature systems PRO-fit (Bruce et al., 2002), E-FIT (Davis et al., 2010) and FACES (Hasel & Wells, 2007), as well as holistic systems EFIT-V (Davis et al., 2010; Valentine et al., 2010) and EvoFIT (Frowd, Bruce, Storås, Spick, & Hancock, 2006). The ACPO (2009) guidelines now require that none of the individual composites are published in the media for identification, but only the morphed composite.

The current thesis considers a further technique that can be applied at the initial interview stage, prior to single composite construction in order to improve composite likeness. This involves the use of reinstating the environmental context in which the (target's/offender's) face had been seen, and is described in more detail below.

Improving facial composites using CR

As discussed above, context, when used as a retrieval cue, significantly increases face-recognition rates (e.g., Evans et al., 2009; Thomson et al., 1982; Wagstaff, 1982). These findings would seem to be applicable to the construction of facial composites. The first published study investigating contextual influence for composite construction was carried out by Davies and Milne (1985). In this experiment, the 'target' was an individual who unexpectedly entered the testing room in search of a calculator. Composites of this person's face were then constructed one week later via Photofit (mechanical, archaic system). This procedure was carried out in the same room for a physical CR (PCR) or in a different room, either following an MCR or no CR (a free face recall). Composite effectiveness was evaluated using a task with composites being matched to a photograph of the target. Results revealed that composites were of better-quality when constructed in the same than a different room; however, those following MCR were of overall best-quality, regardless of the room, leading to an even greater improvement. This could be explained in terms of recall differences between the PCR and MCR. During MRC, individuals were directly instructed to revive memories, including attitudes and feelings. During PCR and free face recall, however, no such instructions were given. Therefore, it would seem reasonable to assume that participants were concentrating on composite construction rather than on the surroundings and contextual information, leading to a qualitative difference between PCR and MCR. When both PCR and MCR were utilised at the same time, composite quality was at its best; however, the improvement from being in the same versus different room was only very small (*MD* = 0.03%), albeit the factor of *room* was still significant, when using an MCR technique. This more informal inspection of their data shows that the PCR is beneficial but not necessary in conjunction with MCR.

Two further unpublished experiments (Ness & Bruce, 2006; Ness, Bruce, & Hancock, 2004) examined CR effects in a different way. Ness et al. (2004) investigated whether reinstating the context by showing video footage (simulated CCTV) of the original target scene would improve the effectiveness of PRO-fit (feature) composites. The first clip shown during initial target viewing was in high quality and 30 seconds long. Participants were unaware that they would subsequently construct a composite of the face. For CR conditions, two clips were made available which were poor quality and 7 seconds long. In one of the clips, the target face was blocked and in the other clip, the face was visible. Results indicated that CR was only successful in increasing composite identification and ratings when participants viewed the video which revealed the target face. This finding could be due to the fact that the background environment was initially not suitably encoded. Although participants were not specifically asked to focus on the target face, they may have done so, especially in the short amount of available time. Also, the length of the CR clips was potentially not sufficient to facilitate memory. In Davies and Milne's (1985) experiment, participants could choose the length of time in which they engaged with the environmental background, whereas in Ness et al.'s (2004) they were constrained to the length of the video. Participants may have spent more time inspecting the target person in the CR video, even if the face was blocked—or maybe *because* it was blocked as this may be considered more distinctive than the environment.

Following this experiment, Ness and Bruce (2006) conducted a second, similar experiment; this time, with the target wearing highly distinctive (bright pink) clothing. Also, there were three different CR conditions: the first was 'no face', in which only the back and side of the head were visible; the second was 'no head' as the face was blocked out; and the third condition showed still photographs of the footage. Participants were now—differently to Ness et al. (2004)—given complete control over the number of times they could view the clips during composite construction. As previously, participant-witnesses were not made aware that they would be asked to construct a composite. Findings revealed that the video footage significantly improved composite likeness, as indicated by ratings and matching tasks. It would have been beneficial to have included a control condition in which no context cues were used at all; the control condition used in this particular experiment still utilised images of the environment, and so it is difficult to assess the extent to which CR improved composite effectiveness.

As research in this area is scarce, the current thesis aims to conduct further, more-intensive research on the impact of both PCR and MCR on the identifiability of facial composites. The ensuing research will utilise modern systems, to be of most use to current policing. Facial composites will be constructed using two types of modern composite systems: the holistic system EvoFIT and the feature system PRO-fit, as both of these system are available to this research project and are widely used within the UK (detailed procedures are described in the next section). Both systems will be used, not necessarily for comparison, but primarily to investigate whether context effects would apply to one or both types. The MCR technique—if shown to be effective—could then potentially be improved further, by combining it with other interviewing techniques already known to facilitate composite construction (e.g., H-CI; Frowd et al., 2008; 2015) or eyewitness memory in general (other mnemonics of the full CI; Fisher & Geiselman, 1992), to be as effective as possible for face construction. A further aim of the thesis is to explore potential reasons for any improvement of the ensuing composites: whether CR facilitates face recognition or face recall, or whether it may be a combination of both.

Detailed procedures of composite systems used in the thesis

The facial composite systems PRO-fit and EvoFIT—both the face-to-face and the online, self-administered version—will be used for the research of the current thesis. These systems differ with regard to the procedure to create a face, and these differences are described in detail in this section. For both systems, participant-witnesses are first interviewed via CI or H-CI, as described above (as well as the CR interview described in more detail in the *Procedure* sections of the Experiments, in Chapter 2) and these interviewing techniques do not differ between systems. Once the interviewing stage is completed, the appropriate system is used, either PRO-fit or EvoFIT as described in the following sections.

Feature system: PRO-fit

Prior to opening the software, a brief description of the process is outlined. For the PRO-fit system, it is explained that participant-witnesses will be presented with a full face, that matches their given description, which can then be further altered by selecting, sizing and positioning different facial features. First, the correct data base is chosen for ethnicity, age and gender. Following this, the participant's face recall that was elicited during the CI or H-CI is inputted in the software. Each facial feature (e.g., hair, eyes, mouth) is selected individually and appropriate characteristics (e.g., hair: short, medium or long) are inserted to reduce the total number of items (approx. 150-200 items per feature) to a more-manageable amount (approx. 20 items) for the witness to view and chose from. If the witness has not provided a sufficient description for a feature, more details are probed during this process.

Once inputted, the initial face is shown, presented in grey-scale (see Figure 1) and it is demonstrated how each feature can be selected, positioned and altered with regard to size, brightness and contrast. Participant-witnesses are then invited to chose which facial feature they would like to work on at each given time. All available items for each particular feature are shown and any close matches can first be 'tagged' before the best match is selected. They are then asked whether any further changes need to be applied to the specific feature in order to create the best possible likeness.

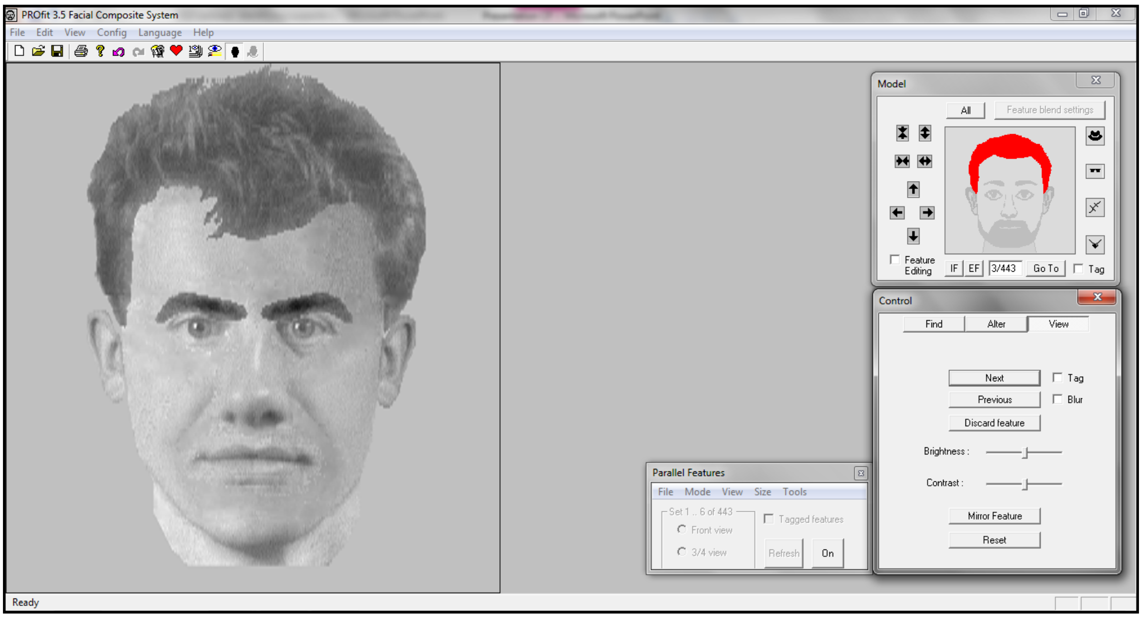

Figure 1. Example PRO-fit screen showing a full face that can be altered by selecting alternative individual facial features.

Once each feature is chosen, a 'Warp' tool can be used to apply changes, such as lowering the hair line or altering features. If necessary, the composite can also be further worked on using a separate software package, for example Photoshop, to add certain details such as moles, wrinkles or hair highlights. Once the witness indicates that the best possible likeness is achieved, the completed composite is saved. Due to emphasis on selecting individual facial features during the construction process, PRO-fit is considered a feature system.

Holistic system: EvoFIT

Face-to-face construction procedure

In contrast to PRO-fit, EvoFIT is considered to be a holistic system as witnesses are shown full faces for selection, displayed in greyscale. A brief overview of how the system works is given at the beginning of the session: it is explained that the witness would be repeatedly shown full internal-feature faces from which to select from and that their selections would be merged to breed further faces for choice; following this initial procedure, they can then use other tools to improve the likeness further, make more precise changes to individual features and add external features (i.e., hair, ears, neck). The appropriate database is then chosen for gender, age and ethnicity and this choice is confirmed by showing the initial screen with arrays of 18 randomly-selected faces. If the witness requests an alternative age or ethnicity, the operator will close the screen and reselect the relevant database.

Once confirmed, the texture of the face is removed, showing internal faces that differ with regard to facial shape (see Figure 2). Using the older selection procedure, the witness is asked to select faces for their *overall* match to the target face (used in Experiment 1 and in the Supplementary experiment of the current thesis); standard procedure now is to select for the *upper face region* (used in Experiments 2-5). The witness is invited to select three faces on the first screen. These three faces are then combined to breed further faces for choice on the second screen (leaving the initial three choices intact). In between each screen, witnesses are prompted by the computer to visualise the original target face. Three further faces are then selected on the second screen, leading to a total of six faces. A third screen is then presented and witnesses are invited to make any alternative choices, whilst maintaining a total of six chosen faces. Now, the unselected faces are removed from the screen, leaving only those six choices intact, and the witness selects one best match: a shape that will subsequently be used to present a variety of facial textures, that is, faces differ with regard to skin tone and shading of individual features.

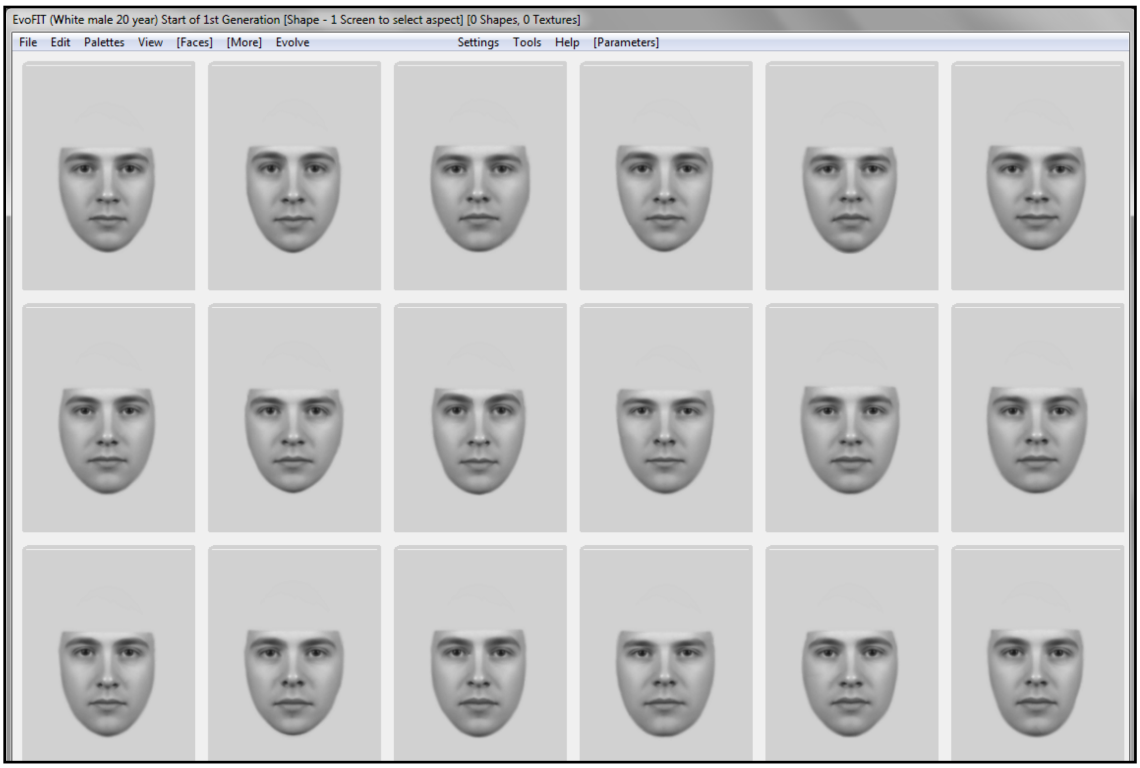

Figure 2. Example EvoFIT screen showing arrays of 18 internal-only faces, here without texture.

The same procedure as for face shape is used to select texture: over the course of three screens a total of six faces are selected. Next, the witness is shown two screens which illustrate arrays of 'combination' faces, faces that have been produced based on the 12 selections made so far (six face shapes and six facial textures). On each screen, one best match is chosen and the operator switches between the two screens so that the witness can indicate the optimal face. Finally, a 'summary' screen is displayed with faces that have been selected previously: faces selected for shape (highlighted with a blue border), texture (green border), both shape and texture (red border) and combination (pink). Witnesses are invited to make a further choice (this is usually the combination face highlighted in pink) and this forms the best face of the first generation.

Following this, the best face is rated on a ten-point Likert scale (1 = poor likeness, 10 = faces are identical); a rating that is included so that the witness reflects upon the likeness but does not influence the generation of face arrays. The entire procedure described above is now repeated for a second time (second generation), except for the difference that initial screens showing face shapes are now also presented with texture. Witnesses again make choices for face shape, texture, combinations and summary screens, whilst visualising the target face between these screens. The best face selected in the second generation is then chosen and the evolving process is ended.

Now, the 'Holistic tool' can be used to make further alterations to the face. Using slider scales, the face can be changed (or remain unaltered) to improve overall characteristics and better-reflect personality traits (scales are presented in the following order: width, age, facial weight, suntan, attractiveness, extraversion, health, honesty, masculinity, threatening, trustworthiness, hardness, dominance, and vertical position of internal features). Texture of the face can also be altered to adjust shading and texture (greyscale level) for eyes, eyebrows, cheekbones, mouth, stubble, moustache, beard, eye bags, laughter lines and deep-set eyes. Once all slides have been used, the resulting face is shown along with the original face (evolved from the second generation). The witness is asked whether the face has been improved or whether they would like to rework the original face using the 'Holistic tool'.

Following this stage, individual facial features can be manipulated in the 'Shape tool' by altering their size and position. Once the witness indicates that the features are as accurate as possible, external features (hair, ears, neck) are added. These are shown in arrays over several screens, where the best match can first be selected and then altered by changing the greyscale colour and brightness of the hair and face as well as reflecting hair horizontally. The likeness can be further improved using a 'Warp tool' (e.g., to lower the hair line; same as in PRO-fit) or Photoshop, as well as by adding hats, extensive facial hair and other accessories. Whenever necessary, tools can be re-used until the best possible likeness is achieved.

Self-administered construction procedure

Additional to the 'face-to-face' EvoFIT system described above, a self-administered version has recently been developed by the EvoFIT production team (www.EvoFIT.co.uk), allowing witnesses and victims to create a composite online in their own homes. This self-administered EvoFIT system (used in Experiments 4 and 5) provides step-by-step screen instructions (see Figure 3) that follow the same construction procedure as used in the traditional, face-to-face method.

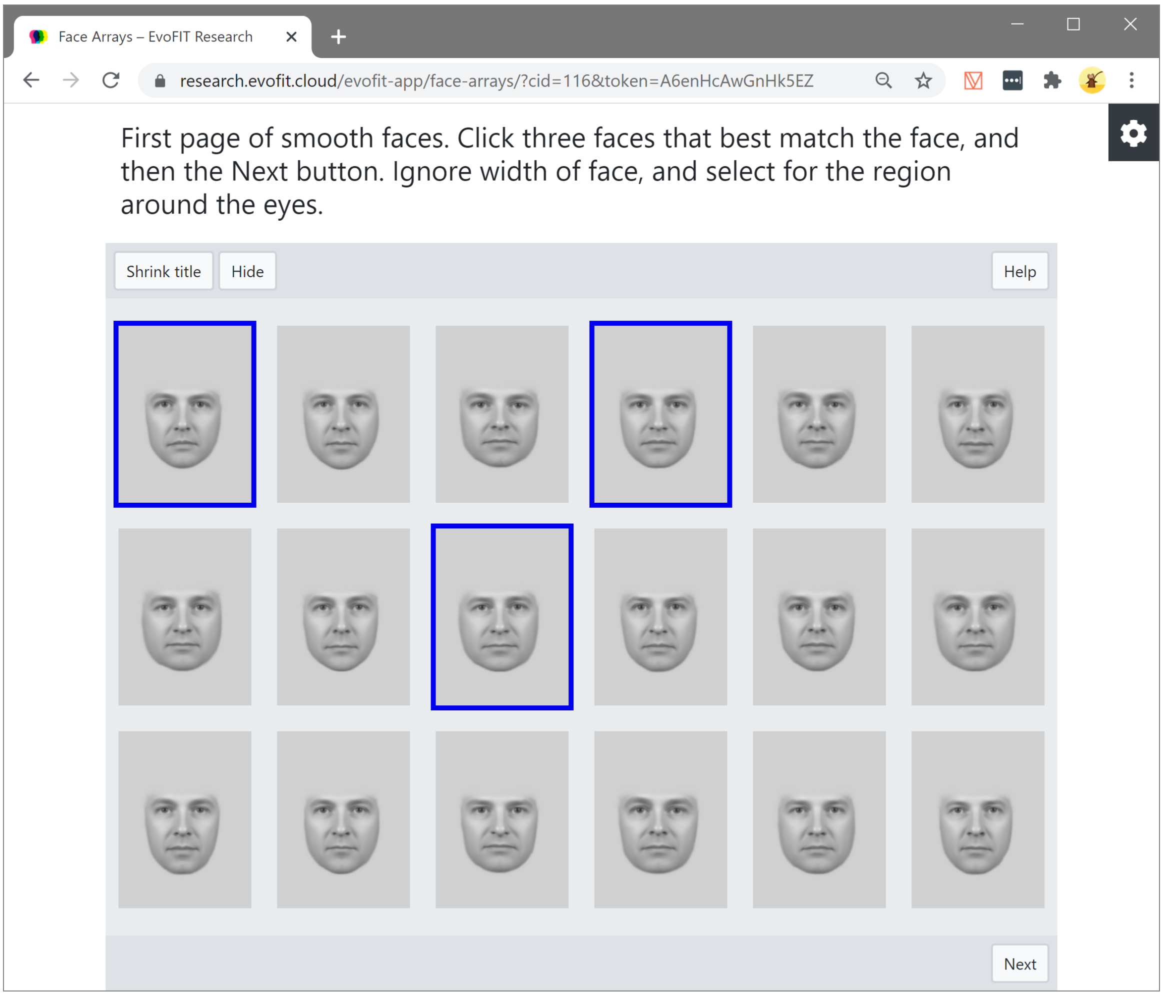

Figure 3. Example EvoFIT online screen showing arrays of internal-only faces, along with instructions.

The next chapter consists of the research paper entitled 'The importance of context reinstatement for the production of identifiable composite faces from memory' that has been submitted for publication and is currently under review. The paper contains all experiments pertaining to the current thesis. Data collection for Experiments 1 to 3 as well as the Supplementary experiment has been completed by the current author. Furthermore—under the author's supervision—, data for Experiment 4 were collected by Ellena Wood and for Experiment 5 by Elizabeth Jackson.

2

THESIS EXPERIMENTS WRITTEN UP AND SUBMITTED AS A RESEARCH PAPER

General introduction

This chapter presents the experimental data from the thesis. These are contained in the following paper, entitled 'The importance of context reinstatement for the production of identifiable composite faces from memory', currently submitted and under revision. Editing changes have been added to to this manuscript, to be consistent with the rest of the thesis.

The importance of context reinstatement for the production of identifiable composite faces from memory

Cristina Fodarella[1*], John E. Marsh[1,2], Simon Chu[1], Palwinder Athwal-Kooner[3], Helen Jones[1], Faye C. Skelton[4], Ellena Wood[1], Elizabeth Jackson[1] and Charlie D. Frowd[1]

[1]School of Psychology, University of Central Lancashire, Preston, UK.

[2]Department of Building, Energy and Environmental Engineering, University of Gävle, Gävle, Sweden.

[3]Department of Psychology, Nottingham Trent University, Nottingham, UK.

[4]School of Applied Sciences, Edinburgh Napier University, Edinburgh, UK.

*Corresponding author: Cristina Fodarella, School of Psychology, University of Central Lancashire, Preston, UK. Email: CFodarella3@uclan.ac.uk

Abstract

Memory is facilitated by reflecting upon, or revisiting, the environment in which information to-be-recalled was encoded. We investigated these 'context reinstatement' (CR) techniques to improve the effectiveness of facial composites—visual likenesses of a perpetrator's face constructed by eyewitnesses. Participant-constructors viewed a face and, after 24-hours, revisited (physical CR) or recalled the environmental context (mental/Detailed CR) before recalling the face and constructing an EvoFIT or PRO-fit composite. Detailed CR increased correct naming and rated likeness of ensuing composites, the size of which was dependent on the extent to which the environment was encoded. Detailed CR was also effective when combined with another interviewing technique (Holistic-Cognitive Interview), with focus on the target's character; it was no more effective when constructors were prompted to engage in greater environmental recall. Analyses indicate that the advantage of Detailed CR was mediated by an increase in face recall. Results should be beneficial to forensic practitioners.

Keywords: Context Reinstatement; Facial composites; EvoFIT; PRO-fit; Holistic-Cognitive Interview

Introduction

Hearing an old song or returning to a place after a long time can unexpectedly revive memories thought to have been forgotten. Any aspect of the environment may trigger a memory and, often, the trigger can be peripheral to the retrieved episode. The phenomenon can be explained by the encoding specificity principle (Tulving & Thomson, 1973) which proposes that encoding a memory involves not only the central aspects of the episode but also information related to the context of the event. Contextual information includes an observer's emotions at the time of encoding and their perception of the environment, which can involve a range of senses (e.g., visual, auditory, olfactory). For the visual modality in particular, there is usually a large amount of information that a person perceives about objects in the visual scene in terms of their shape, size and colour as well as their spatial arrangement and dynamics (i.e., whether and how they move). Such contextual information may later potentially act as retrieval cues, facilitating access to the desired (or 'target') memory (e.g., returning to a childhood playground, a long-forgotten conversation, a person's facial appearance). Recalling these ancillary retrieval cues should promote recall of the target memory.

The benefit of contextual reinstatement (CR) has been repeatedly demonstrated in the literature (e.g., see Smith, 2013 for a review; Smith & Vela, 2001 for meta-analysis), and in a variety of different areas including verbal memory (e.g., Campeanu, Craik, Backer, & Alain, 2014; Godden & Baddeley, 1975; Smith, 1979), object memory (Barak, Vakil, & Levy, 2013; Koen, Aly, Wang, & Yonelinas, 2013; Levy, Rabinyan, & Vakil, 2008), eyewitness memory (e.g., Dando, 2013; Wagstaff, Wheatcroft, Caddick, Kirby, & Lamont, 2011; Wong & Read, 2011) and facial memory (e.g., Davies & Milne, 1985; Shapiro & Penrod, 1986). Arguably one of the most well-known verbal learning studies was conducted by Godden and Baddeley (1975) in which divers learned word lists either on land or under water. Considerably more words (46%) were retrieved from memory if recalled in the same environment as encoding, than in the alternate environment, suggesting that reinstating the environmental context during retrieval facilitates memory recall.

Rather than reinstating the environmental context by physically returning to the location of encoding, comparable benefits to memory may be achieved by mentally visualising and recalling the encoding context prior to retrieval (e.g., Smith, 1979), a cognitive or 'mental' context reinstatement (MCR). In addition, the psychological state of the observer forms retrieval cues that can also facilitate memory recall. When a specific mood (Bower, 1981; Bower & Cohen, 1982; Eich & Metcalfe, 1989) or level of arousal (Clark, Milberg, & Ross, 1983) is congruent at learning and test, memory recall can be facilitated. Similarly, reproducing the pharmacological state of participants at learning (e.g., sleep deprivation, or influence of drugs or alcohol) can help trigger memory during recall, an effect termed state-dependent learning (see Eich, 1980; Eich, Weingartner, Stillman, & Gillin, 1975). This indicates that besides physical factors, other associative and cognitive elements also form retrieval cues (Anderson & Bower, 1972).

While research assessing the effect of CR on the recall of verbal memory is theoretically interesting and may be applicable to a real-world setting, methodologies used in the literature often lack ecological validity. Remembering individual words through reinstating the context in form of a background image (e.g., Smith & Manzano, 2010) or the position on a screen (e.g., Macken, 2002; Murnane & Phelps, 1993) is not necessarily indicative of how memory might be improved in everyday situations. However, some research has also been *applied* in nature, most prominently on eyewitness memory. Results in this domain also favour the use of contextual cues for increasing recall (Dando, Wilcock, Milne, & Henry, 2009; Hammond, Wagstaff, & Cole, 2006; Wagstaff, Cole, Wheatcroft, Marshall, & Barsby, 2007; Wong & Read, 2011), with little or no increase in information that is inaccurate (e.g., Davis, McMahon, & Greenwood, 2005; Emmett, Clifford, & Gwyer, 2003; Memon & Bruce, 1995; Milne & Bull, 2002). One of the best-known methods for facilitating recall is the Cognitive Interview (CI).

The CI is an interviewing procedure usually administered by police officers for eyewitnesses to recall information about a crime (e.g., Geiselman, Fisher, MacKinnon, & Holland, 1985; Wells, Memon, & Penrod, 2007). It consists of a series of techniques or *mnemonics*, such as to report everything (to try to prevent witnesses from holding back information) and to attempt recall in a different temporal order (to encourage use of different retrieval paths). While the CI has been extensively assessed and improved (e.g., for a review, see Frowd, 2011), one of the mnemonics that has been consistently included since the original interview is to reinstate the context. Using Mental Context Reinstatement (MCR), observers are asked to mentally reinstate the environmental context at the point of encoding, taking into account other physical conditions (e.g., smells, sounds) as well as their own psychological states (e.g., reactions, mood). MCR has been consistently shown to lead to the retrieval of a greater amount of information than a standard ('question and answer' type) police interview with only minor (or no) increase in inaccurate information recalled (Memon, Meissner, & Fraser, 2010). In fact, as well as the 'report everything' mnemonic, the MCR is considered to be the most effective mnemonic for triggering retrieval (Davis et al., 2005; Milne & Bull, 2002).

Context can also facilitate memory for faces (e.g., Brown, 2003; Rainis, 2001; Shapiro & Penrod, 1986; Thomson, Robertson, & Vogt, 1982). Studies using a line-up (or an identity parade) show that participants tend to recognise a target face significantly more accurately when tested in the same room where facial encoding initially occurred rather than in a different room (Evans, Marcon, & Meissner, 2009; Wagstaff, 1982; Wong & Read, 2011). MCR is also effective in increasing correct recognition rates (Cutler, Penrod, & Martens, 1987; Malpass & Devine, 1981).

Another practical aspect of facilitating memory for faces, also likely to benefit from context reinstatement, be it physical or mental, is the construction of facial composites. Composites are facial likenesses produced from witnesses' and victims' memory of a perpetrator of crime. These visual representations of a face are used by law enforcement to identify potential suspects. There are many documented cases where facial composites have led to a serious criminal (e.g., a rapist, murderer, confidence) being identified and later—following further compelling evidence—convicted (e.g., Frowd, Pitchford, et al., 2012), and so methods to improve their effectiveness are both theoretically interesting and valuable to security.

Sketch artists have traditionally worked with eyewitnesses to sketch a face, but production systems have since been developed, initially from mechanical 'feature' types (e.g., Photofit and Identikit), which are now archaic, to computerised 'feature' systems (e.g., PRO-fit, E-FIT, FACES), and, more recently, 'holistic' or recognition systems (e.g., EvoFIT, EFIT-V [now called EFIT-6], ID). A detailed review of the systems can be found in Frowd (2017). In essence, feature systems allow an eyewitness to select individual facial features (e.g., eyes, nose, mouth, hair) to construct a face, while holistic systems involve witnesses repeatedly selecting from screens of whole faces (or whole-face regions), with choices combined to 'evolve' a face. Composite recognition then occurs based on both their featural information, that is, individual facial features, and their configural information, that is, the spacing between these features (see Frowd et al., 2014). Research suggests that holistic systems are more effective at producing identifiable composites presumably because they are based on face recognition processes (which are more stable over time; Davies, 1983) rather than recall processes (e.g., Frowd, 2017; Frowd et al., 2010, 2015).

Davies and Milne (1985) appears to be the first published study to investigate the influence of context on facial-composite construction. In their work, one of four target individuals was seen entering a testing room and searching for a calculator. After a one-week delay, observers constructed a composite face using the now archaic Photofit while in the same room or a different room, and following a guided memory procedure for recalling the environmental context or without such a procedure. The ensuing composites were given to other people to match to a recent photograph of the targets. Matching scores indicated that significantly better-quality composites were produced (i) under guided memory (cf. spontaneous memory recall) and, to a lesser extent, (ii) in the same (cf. different) room. In other work, Ness and Bruce (2006) investigated a novel procedure for reinstating physical context for face construction. Using the modern PRO-fit system, constructors who were given the opportunity to review video footage of the encoding environment (for a target identity just seen) created composites with better likenesses, faces that were matched more accurately to identity.

The current project investigates if cues available in the environment would enable participants to create a more accurate face using modern composite systems after a realistic delay. In Experiment 1, participants viewed an unfamiliar target face and underwent one of three procedures the following day to reinstate context. Participants were met by an experimenter either in the original environment where target encoding had taken place (Physical CR), or in a different environment for a mental CR. For the latter, participants underwent either (i) Minimal CR, where they were instructed to "think back" to the environment, or (ii) Detailed CR, consisting of recall of both the environment and the person's mood and feelings at the time[3]. Afterwards, participants freely described the face (using further mnemonics of the CI) and constructed a single composite of it using either a typical feature system (PRO-fit) or a typical holistic system (EvoFIT). The resulting composites were given to other people to assess for effectiveness by trying to name them or by providing ratings of 'goodness of fit' (likeness).

Based on the aforementioned research, it was predicted that Detailed and Physical CR would improve a constructor's face-recognition ability during composite construction for both systems. Therefore, more identifiable composites would be constructed (i) under Detailed CR than Minimal CR, and (ii) under Physical CR than Minimal CR. The literature also suggests that composites should be constructed more effectively from holistic than from feature systems, and we expected to observe the same result.

---

[3] Note that '*Detailed CR*' is equivalent to the type of elaborative Mental Context Reinstatement as used in other studies of memory (but not facial composite construction). Since a simple "think back" type (Minimal) CR is already in use with face construction, we draw a distinction between Minimal and Detailed CR.

Davies and Milne (1985) noted that there appears to be two main mechanisms by which face construction could be rendered more effective using reinstatement techniques. One mechanism could be that CR increases witnesses' face *recognition*, leading to more accurate selection of individual facial features or, in the case of a holistic system, whole-face regions. The other mechanism, which does not preclude the former, could be that CR promotes a better memory of the face. This explanation can be evidenced by an increase in witness *recall* of the face and may allow witnesses to construct composites with more accurate detail—for example, to create faces with more accurate feature shapes. One aim of the current work, should an advantage of context reinstatement be replicated, is to provide an indication of the likely engendering mechanism. Given the importance of feature information involved in face construction using Photofit and other feature systems (e.g., Frowd, 2015, 2017), our initial hypothesis was that benefit would be a function of improved memory ability, as evidenced by greater recall of the face. Davies and Milne (1985) note that the effect of both processes in their work may have been hampered by task difficulty and low experimental power, so here we attempt to overcome these issues by combining face recall over a number of studies (Meta-analysis) and changing the timing of recall mnemonics (Experiment 5).

EXPERIMENT 1: Using context reinstatement to facilitate face construction using modern composite systems

Method

The most effective facial composite research mirrors, to the greatest extent possible, the real-life situation in which a witness or victim observes a (usually unknown) perpetrator during a crime, and is required to create a visual likeness of the face after a minimum of 24 hours (Frowd et al., 2005). When the composite is subsequently shown to police officers or the public, anyone who is familiar with the individual may recognise the composite and be able to provide investigators with a possible identity. In this experiment, we model this situation in two stages: composite construction, where a composite of a target face is created from memory nominally 24 hours later, by someone unfamiliar with the target, and composite naming, where someone familiar with the target is asked to identify the composite. We also collect ratings of similarity (likeness) between each composite and its target face as a supplementary measure of composite utility. Different groups of participants take part in each of these three stages.

In the experiments presented here, we chose to present target faces as static images; this is usually the case in composite research (e.g., Fodarella et al., 2017; Frowd, Skelton, et al., 2012; Gawrylowicz, Gabbert, Carson, Lindsay, & Hancock, 2012; Hasel & Wells, 2007; Kehn, Renken, Gray, & Nunez, 2014). Although it could be argued that a staged event or use of video stimuli would be more realistic, composite identifiability changes little following a target presented in video or as a static image (Frowd et al., 2015). Therefore, use of static images seems to generate realistic findings, generalisable to viewing a face in motion.

Participants viewed a target face in the knowledge that they would produce a composite on the following day. This tends to promote an intentional type of encoding likely to increase memory (Shapiro & Penrod, 1986); however, this design choice is not different to how real eyewitnesses may remember faces. In many cases, witnesses and victims make a deliberate attempt to remember the appearance of an offender's face, having the intuition that they will be asked about the face at a later date (Frowd et al., 2015), so we copy this method of encoding here.

The aim was for the experiments to have sufficient power to be able to detect a medium-to-large effect size, should one exist. While dependent on variability of data, this effect size usually leads to around (at least) 50% difference in mean correct naming—for example, a mean of 20% correct in one condition and 30% in another [(30 – 20) / 20 = 50%]. This aim was achieved using ten target identities and at least 18 participants per group for composite naming (Frowd, 2015). Such an increase should translate into a useful benefit for policing; in the paper, we report Cohen's *d* for composite naming, where a value of 0.5 is considered a 'medium' effect and 0.8 as 'large' (Cohen, 1988).

Stage 1: Composite construction

Design

A 2 (System: EvoFIT, PRO-fit) × 3 (CR: Minimal, Physical, Detailed) between-participants design was used for composite construction. Context reinstatement was manipulated over three conditions. The Minimal CR ('control') condition consisted of 'thinking back' to the encoding environment, with face recall and composite construction then taking place in a different environment to that in which the face had been seen (encoded). The Physical CR condition was the same as the first except that face recall and composite construction were conducted in the same environment (room) as encoding. The third condition was Detailed CR. As the first condition, a different environment was used to that in which the face had been encoded, and, prior to face recall and composite construction, participants were asked to recall both the environment and their moods and feelings at that time. In all cases, after the relevant CR manipulation and face recall, each participant created a single composite using either the holistic system EvoFIT or the feature system PRO-fit.

Participants

An opportunity sample of 60 (24 males, 36 females; M_{age} = 30.3, SD_{age} = 11.3 years) participants took part on a voluntary basis. They were staff and students from the University of Central Lancashire (UCLan). To simulate the usual situation for real eyewitnesses, participants were recruited on the basis of being unfamiliar both with the testing environment (a student café) and with the target faces.

Materials

Target faces were 10 current characters from a popular UK soap opera, Coronation Street, sourced from the Internet (Ken Barlow, Leanne Battersby, Peter Barlow, Michelle Connor, Jason Grimshaw, Tracey McDonald, David Platt, Kirk Sutherland, Sally Webster and Sophie Webster). These pictures were good quality, shown in full-frontal pose with minimal facial expression; male actors had little or no facial hair. Stimuli were printed in colour to dimensions of 8 cm (width) × 10 cm (high).

Procedure

Participants were tested individually, and the procedure was self-paced. To allow good control of exposure to the testing environment (described below), each person was met at a convenient meeting point and taken to the room used for target encoding. Participants were randomly assigned to one of the six experimental conditions, defined above, with equal sampling. Each person was shown a target picture, randomly selected, and asked whether the identity was familiar; if it was, another picture was similarly shown. For the first face reported to be unknown, participants were given 60 seconds to remember the face. For this part of the procedure, the participant was aware that a composite would be constructed of this face the following day. The experimenter was blind to the identities included in the experiment as well as the face seen by each participant. Participants viewed the face in a student café located in an unfamiliar building on the UCLan Preston campus, an environment selected to be unfamiliar to participants as well as rich in environmental recall cues: it included tables, chairs, a television, plants, a vending machine and a small counter selling refreshments and confectionary. Participants' unfamiliarity with the environment was established by asking whether they had visited the café prior to the experiment. If anyone had reported previously visiting the café, they would have been excluded (there were no such occurrences). The experimenter made no reference to the importance of the café (to allow environmental context to be encoded incidentally).

Following a delay of 20 to 28 hours, according to assignment, participants met the researcher either in the student café for Physical CR, or in a neutral office space for Minimal and Detailed CR. For Minimal and Physical CR, participants were first interviewed via Cognitive Interview[4] in which they were asked to mentally

[4] Please note that the term 'Cognitive Interview' within facial composite research differs from the full 'Cognitive Interview' (Fisher & Geiselman, 1992). In relation to composites, the CI is a more concise version and usually only includes mnemonics for rapport building, visualisation and free recall of the target face (see Frowd, 2011 for an in-depth review of the CI for facial-composite construction). In the current paper, we refer to CI (and in Experiment 2 to 'Holistic-Cognitive Interview') as used within composite research.

visualise the target face and then freely recall it in as much detail as possible. In the Detailed CR condition, prior to face recall, participants were also asked to mentally visualise the environment in which they saw the target face, and then to reflect silently on their mood and psychological state at the time of viewing. Following this, participants were asked to freely recall the environment as well as their mood and feelings at the time. As elsewhere, participants then freely recalled the face.

Once face recall had been completed, each participant constructed a single composite of the target on a laptop computer using EvoFIT or PRO-fit. The experimenter controlled the relevant software program and took the participant through the procedure to construct the face, the aim of which was to construct the best likeness possible. A detailed description of the relevant procedure for each system can be found in Fodarella, Kuivaniemi-Smith, Gawrylowicz, and Frowd (2015). In brief, for PRO-fit, participants were asked to select the best matching facial features (eyes, brows, nose, mouth, hair, ears) for their given target, and then resize and position each feature to give the best likeness possible. For EvoFIT, participants were asked to repeatedly select overall best matches from arrays of internal features to evolve a face, use software tools to enhance the overall likeness and facial features, and then add external features (hair, ears, neck).

The procedure to construct a composite face took between 20 and 45 minutes per person including debriefing.

Stage 2: Composite naming

Design

This time participants who were familiar with Coronation Street characters were recruited. They were given a set of composites to name that had been constructed using one of the six individual procedures in Stage 1. Thus, the design was a 2 (System: EvoFIT, PRO-fit) × 3 (CR: Minimal, Physical, Detailed) between-participants. It is worth mentioning that this design may lead to an elimination strategy, with participants deciding between possible identities when attempting to name a face. In real life, an elimination strategy may also be involved to some extent, as other information (e.g., offender's build, age and accent) is usually published alongside the composite. In addition, given the nature of the design, one would imagine that any such strategy would apply equally to each experimental condition. Overall, the naming procedure used here has been found to lead to consistent results (e.g., Frowd et al., 2015), and to be a good indicator of identification of composites in the real world (Frowd, Pitchford, et al., 2012).

Participants

Forty-eight (42 males, 6 females; M_{age} = 41.0, SD_{age} = 15.3 years) volunteer Coronation Street fans were recruited outside the 'Coronation Street Tour Set' in Manchester. These participants were familiar with the relevant identities, and the testing procedure (as detailed in *Procedure* below) ensured that they would recognise a minimum of 80% of the targets.

Materials

Sixty composites and the 10 target photographs were standardised to dimensions of 8 cm (width) × 10 cm (high) and printed individually in greyscale; see Figure 1 for example composites.

Figure 1. Example EvoFITs (top row) and PRO-fits (bottom row) of Coronation Street character Leanne Battersby constructed following (a) Minimal, (b) Physical, (c) Detailed CR. Each composite face was constructed from memory by a different person. Due to copyright issues, we are unable to reproduce the target face used in the experiment, but an Internet search could readily reveal the appearance of this actress, Jane Danson.

Procedure

Participants were randomly allocated with equal sampling to the two experimental factors (i.e., one of six individual conditions) described above. During briefing, they were informed that they would view and attempt to identify a set of composites depicting Coronation Street characters. Once consented, participants were shown 10 facial composites (based on assignment) sequentially to name. Participants were encouraged to guess, and it was explained that it was also acceptable to give identifying semantic information if they were unable to remember the name, or not to give a name at all. Once all 10 composites had been presented, the target photographs were shown likewise for naming. We applied an *a priori* rule, to ensure participants were suitably familiar with the relevant identities, such that each person was required to correctly name a minimum of eight of the ten targets (or another person was to be recruited as replacement); as it turned out, all participants met this rule. Participants each received a different random order of presentation for composites and target photographs. Testing sessions lasted for approximately 10-15 minutes, including debriefing.

Stage 3: Composite ratings of likeness

Design

As a supplementary measure of composite quality, a third group of participants rated the likeness of each composite against the relevant target photograph. The design was within-participants for the two experimental factors, System and CR.

Participants

Eighteen (8 males, 10 females; M_{age} = 32.3, SD_{age} = 12.6 years) UCLan student volunteers were recruited on the basis of being unfamiliar with the Coronation Street TV soap.

Materials

The 60 composites were printed alongside the relevant target photograph, in greyscale, to dimensions of approximately 6 cm (width) × 8 cm (high).

Procedure

Participants were tested individually. They were shown all sixty composites alongside the associated target photograph, sequentially, and were asked to rate the overall likeness of each on a scale from 1 (poor likeness) to 7 (good likeness). The order of presentation of composites was randomised for each person. The task was self-paced and took approximately 15 to 20 minutes, including debriefing.

Results

Composite naming: By-participants analysis

Participant responses to target photographs and facial composites were scored for accuracy: a value of 1 was assigned when a given response was correct and 0 otherwise (incorrect name or no name given). Participants correctly named all of the target photographs and so familiarity with the relevant identities was at 100.0%. This result suggests that all of the composites had the potential to be correctly named by all of the participants. As can be seen in Table 1, mean correct naming of composites was considerably less than 100%, which is the usual case for this type of error-prone facial stimuli.

Table 1. Percentage of EvoFIT and PRO-fit composites correctly named by Context Reinstatement

	Context Reinstatement (CR)†			
*System**	Minimal	Physical	Detailed	Mean
EvoFIT	23.8 (17.7)	22.5 (14.9)	33.8 (9.2)	26.7 (14.6)
PRO-fit	6.3 (5.2)	7.5 (4.6)	13.8 (9.2)	9.2 (4.0)
Mean	15.0 (15.5)	15.0 (13.2)	23.8 (13.6)	17.9 (14.4)

Note: † Significant main effect of CR, $p < .05$; * Significant main effect of facial-composite System, $p < .001$. In parentheses are (by-participant) SD values.

The number of correct responses per participant was analysed using Independent Samples ANOVA for CR type (Minimal, Physical, Detailed) and System (EvoFIT, PRO-fit). There was a significant main effect of CR [$F(2,42) = 3.27$, $p = .048$, $\eta_p^2 = .14$] and two-tailed Simple Contrasts indicated that Detailed promoted significantly higher naming rates than both Minimal (*p* = .032, Cohen's *d* = 0.66) and Physical (*p* = .032, *d* = 0.66) CR; an additional t-test revealed that there was no significant difference between Minimal and Physical [$t(30) < 0.001$, $p = 1.00$, $d < 0.01$] CR. The main effect of System was also significant [$F(1,42) = 29.40$, $p < .001$, $\eta_p^2 = .41$], with EvoFIT composites named significantly higher than PRO-fit composites. The interaction between CR and System was not significant [$F(2,42) = 0.20$, $p = .82$, $\eta_p^2 = .01$].

Composite naming: By-participants and by-items analysis

The data was additionally analysed using Generalised Estimating Equations (GEE) which is a type of regression analysis combining both by-participants and by-items analysis. Context and System were entered as predictors, and the response variable (DV) was composite naming accuracy, a binary response. The Working Correlation Matrix was set to 'exchangeable', and the best estimator became apparent as model-based. As the interaction between Context and System was non-significant (*p* = .79), it was removed accordingly from the model. Subsequently, the new model [*QIC* = 426.00, *QICC* = 427.74] emerged significant for Context [$X^2(2) = 7.95$, $p = .019$]. Detailed led to better-named composites compared to both Minimal [$B = 0.60$, $SE(B) = 0.25$, $X^2(1) = 5.64$, $p = .018$, $Exp(B) = 1.83$, *95%CI* (1.11, 3.00)] and Physical [$B = -0.60$, $SE(B) = 0.25$, $X^2(1) = 5.64$, $p = .018$, $Exp(B) = 0.55$, *95%CI* (0.33, 0.90)]. However, Physical and Minimal did not significantly differ [$B = 0.00$, $SE(B) = 0.27$, $X^2(1) < .001$, $p = 1.00$, $Exp(B) = 1.00$, *95%CI* (0.59, 1.71)]. Further, System appeared as a significant predictor [$X^2(1) = 32.11$, $p < .001$], with EvoFIT composites being better-named than PRO-fit composites [$B = -1.30$, $SE(B) = 0.23$, $X^2(1) = 32.11$, $p < .001$, $Exp(B) = 0.27$, *95%CI* (0.17, 0.43)].

Composite likeness rating

A pattern of results generated by Composite naming was similarly observed in the Composite likeness rating scores (Table 2). Participant mean ratings were analysed by Repeated Measures ANOVA for CR (Minimal, Physical, Detailed) and System (EvoFIT, PRO-fit). There was a significant main effect of CR [$F(2,34) = 430.06$, $p < .001$, $\eta_p^2 = .96$], and two-tailed Simple Contrasts revealed that composites from Detailed CR were rated as significantly higher likenesses than both Physical ($p < .001$) and Minimal ($p < .001$); this time, though, a t-test indicated an advantage of Physical over Minimal [$t(17) = 6.52$, $p < .001$]. There was also a significant main effect of System [$F(1,17) = 535.24$, $p < .001$, $\eta_p^2 = .97$], with EvoFIT composites rated as better likenesses than PRO-fit composites.

The interaction between system and CR was also significant [$F(2,34) = 17.16$, $p < .001$, $\eta_p^2 = .50$]. In this analysis, all pairwise comparisons were reliable ($p < .001$), and so we suspected that the interaction was being driven by a larger effect for System or for CR. We calculated mean difference (*MD*) scores for Minimal and Physical CR, and Physical and Detailed CR for both systems. Using these *MD* scores, while there was no significant difference between systems for Minimal and Physical [$t(17) = 0.75$, $p = .46$], Physical and Detailed was significantly greater for EvoFIT than for PRO-fit [$t(17) = 4.18$, $p = .001$]. The latter result suggests that the improved effectiveness of Detailed CR compared to Physical CR was greater for composites created using EvoFIT than PRO-fit.

Table 2. Mean likeness ratings (SD) for EvoFIT and PRO-fit composites by Context Reinstatement

	*Context reinstatement (CR)**			
System†	Minimal	Physical	Detailed	Mean
EvoFIT	2.5 (0.3)	2.9 (0.2)	4.3 (0.3)	3.2 (1.0)
PRO-fit	1.7 (0.2)	2.0 (0.2)	2.9 (0.2)	2.2 (0.6)
Mean	2.1 (0.6)	2.4 (0.6)	3.6 (1.0)	2.7 (0.9)

Note: † Significant main effect of System, $p < .001$; * Significant main effect of CR, $p < .001$; Significant CR × System interaction, $p = .001$. The likeness rating scale was from 1 (poor likeness) to 7 (good likeness).

Additional analyses

There are other methods of assessment that can be carried out on data produced from the experiment. First, incorrect (mistaken) naming of composites can be analysed to give a measure of composite misidentification and assessment of response bias (guessing). Second, information recalled about the face from each participant who constructed a composite can also be assessed, hypothesised to be greater under Physical and Detailed CR (cf. Minimal). In order to maintain brevity in the paper and due to issues of experimental power, these additional assessments are conducted later in a separate section on meta-analyses and additional assessments. In brief, there was no reliable difference by CR in terms of incorrect names given, and significantly greater information was recalled overall about the target face by participant constructors following Detailed compared to Minimal CR.

Discussion

The naming and likeness rating data taken together replicate Davies and Milne's (1985) findings: composites were most effective in the Detailed CR condition, followed by Physical CR, with the control condition (Minimal CR) leading to the least effective composites. This was shown using two types of composite system: a modern feature system, PRO-fit, and one of the newer recognition type systems, EvoFIT, with the latter system outperforming the former. While correct naming scores did not reveal a significant benefit for Physical (cf. Minimal) CR, this was apparent in the likeness rating data (thus also supporting the previous research by Ness & Bruce, 2006); the rating data also revealed that the improvement from Physical to Detailed CR was stronger for faces constructed using EvoFIT than PRO-fit. Overall, the data suggest that detailed recall of the physical and psychological context is advantageous for reproducing faces from memory using two contrasting methods of face production. Our working hypothesis is that Detailed CR would be effective as it improved a constructor's memory of the face, allowing them to more-effectively process the face as a whole—that is, leading to a better end result whether that is achieved through selection of individual features (PRO-fit) or from face arrays (EvoFIT).

Experiment 2 attempted to replicate the effect of the Detailed CR procedure and ascertain whether participants would be able to construct even more effective composites using an enhanced type of interview. In Experiment 1, prior to face construction, witnesses were interviewed using a Cognitive Interview (CI) to help them recall the appearance of the target face. An enhancement of this interview, termed the Holistic-Cognitive Interview (H-CI), involves focus on the personality or character of the face (e.g., Frowd, Bruce, Smith & Hancock, 2008; Frowd, Nelson, et al., 2012). Specifically, after face recall, witnesses are asked to reflect silently on the personality of the face for 60 seconds and then make seven personality judgements (e.g., intelligence, friendliness, kindness) on a three-point Likert scale. The procedure is believed to improve a constructor's face recognition by encouraging holistic processing of the face. The resulting composite is more identifiable than that produced following the more standard face recall CI (see Frowd et al., 2015 for a meta-analysis).

Therefore Experiment 2 involves three factors: CR (Minimal CR, Detailed CR), method of witness interview (CI, H-CI) and system (EvoFIT, PRO-fit). If Detailed CR promotes a better memory of the face, then using H-CI should improve composite effectiveness even further. Based on the Experiment 1 composite naming data, Physical CR was not considered any further since there was no evidence that it helped participants to construct a more identifiable composite; we consider potential explanations for this (null) effect in the General Discussion.

EXPERIMENT 2: Combining Context Reinstatement and Holistic-Cognitive Interview

Stage 1: Composite construction

Design

A 2 (System: EvoFIT; PRO-fit) × 2 (CR: Minimal, Detailed) × 2 (Interview: CI, H-CI) between-participants design was used. Thus, nominally 24 hours after encoding an unfamiliar target face in an unfamiliar environment (café), participants underwent Minimal or Detailed CR, described the face using CI or H-CI and constructed a single composite using EvoFIT or PRO-fit.

There was also a change in the procedure used with EvoFIT. Recent findings reveal that composites are produced more identifiably if constructors are requested to select a match for the *upper half* rather than for the overall match of the presented internal features arrays when evolving the face (Fodarella et al., 2017). This new procedure tends to produce composites with a more accurate upper facial region, an area known to be important to recognition of both facial photographs (Goldstein & Mackenberg, 1966; Pellicano, Rhodes, & Peters, 2006) and facial composites (Laughery, Duval, & Wogalter, 1986). This instruction is now used regularly with witnesses and victims of crime (Frowd, Portch, Killeen, Mullen, Martin, & Hancock, 2019), and so including it here allows results to reflect current forensic practice.

Participants

Sixty-four (26 males, 38 females; M_{age} = 29.1, SD_{age} = 8.6 years) UCLan staff and students took part voluntarily. As in Experiment 1, participants were recruited on the basis of being unfamiliar with the target identities.

Materials

To have greater confidence in the generalisability of the CR advantage, target faces were drawn from a completely different pool of identities, current football players who play at international level in the UK. Recruitment of participants followed the same criteria as before (i.e., participants were recruited to be unfamiliar with targets for the face construction stage, but familiar for the naming stage). Target faces were eight photographs (Ross Barkley, Gary Cahill, Michael Carrick, Joe Hart, Harry Kane, Adam Lallana, James Milner and Jack Wilshere) sourced from the Internet and of the same standard as Experiment 1, printed in colour (8 cm × 10 cm).

Procedure

Each participant viewed a target face in the same café as in Experiment 1, and met with the experimenter the following day in a different room to construct a composite of this face. The procedure was the same as in Experiment 1, except for the following differences. At the start of the second session, as Physical CR was no longer used, participants engaged in either Minimal CR or a Detailed CR. After this CR manipulation, participants were interviewed using either (i) the CI, in which participants were asked to mentally visualise and then freely recall the target face in as much detail as possible, or (ii) the H-CI, as (i) but then were asked to reflect silently on the personality of the face for 60 seconds and make seven personality attributions, rating each on a three-point Likert scale. After the interview, participants created a single composite using either PRO-fit, as described in Experiment 1, or EvoFIT. The EvoFIT procedure differed from that used in Experiment 1, in so far as participants were instructed to select best matches in the presented arrays for the upper facial region; after evolving the face, participants were requested (as before) to focus on all aspects of the face (not just the upper region), in order to enhance the facial appearance using the software tools.

Stage 2: Composite naming

Design, Material and Procedure

Composites were named by a different group of participants using the same three-factor design as described in Stage 1. Materials were 64 composites and eight target photographs, printed individually in greyscale as in Experiment 1. The procedure used to name the composites was also the same as in Experiment 1, except that there were now 64 composites and eight targets, and participants were randomly allocated, with equal sampling, to the eight cells of the design.

Participants

Eighty (77 males, 3 females; M_{age} = 40.6, SD_{age} = 14.5 years) participants took part on a voluntary basis. They were recruited opportunistically from Manchester Football Museum on the basis of being familiar with the target identities.

Stage 3: Composite rating

Design, Materials and Procedure

The design for obtaining composite likeness ratings was the same as in Stage 1 but here, as before (in Experiment 1), was within-participants. Materials were 64 composites and eight target photographs, printed as in Experiment 1. The procedure was also the same as in Experiment 1, except that participants rated 64 composites against their associated target photographs.

Participants

Eighteen (7 males, 11 females; M_{age} = 34.2, SD_{age} = 11.8 years) UCLan staff and students volunteered and were recruited opportunistically. Participants were recruited to be unfamiliar with the target identities.

Results

Composite naming: By-participants analysis

Participant responses to targets and composites were scored for accuracy. Familiarity with the target identities was at 100.0%; a summary of composite naming is shown in Table 3.

Table 3. Percentage of EvoFIT and PRO-fit composites correctly named by Context Reinstatement and Interview

	*Context Reinstatement**				
	Minimal		Detailed		
System†	CI	H-CI‡	CI	H-CI‡	Mean
EvoFIT	10.0 (9.9)	22.5 (17.5)	26.3 (16.1)	31.3 (18.9)	22.5 (17.3)
PRO-fit	2.5 (5.3)	6.3 (6.6)	5.0 (6.5)	6.3 (8.8)	5.0 (6.8)
Mean	10.3 (12.9)		17.2 (17.6)		13.8 (15.7)

Note: † Significant main effect of System, $p < .01$; * Significant main effect of CR, $p < .05$; ‡ Significant main effect of Interview, $p < .05$; Significant CR × System interaction, $p < .05$. In parentheses are (by-participant) SD values.

Independent Samples ANOVA revealed a significant main effect of (i) CR [$F(1,72) = 6.26$, $p = .015$, $\eta_p^2 = .08$], indicating Detailed CR led to better-named composites than Minimal CR, (ii) Interview [$F(1,72) = 4.19$, $p = .044$, $\eta_p^2 = .06$], with H-CI ($M = 16.6$, $SD = 17.3$) leading to better-named composites than CI ($M = 10.9$, $SD = 13.6$), and (iii) System [$F(1,72) = 40.55$, $p < .001$, $\eta_p^2 = .36$], as EvoFIT composites were named more accurately than PRO-fit composites. The interaction between CR and System was also significant [$F(1,72) = 4.19$, $p = .044$, $\eta_p^2 = .06$], as Detailed CR significantly increased naming (cf. Minimal CR) for EvoFIT [$t(38) = 2.43$, $p = .020$,

d = 0.66] but not for PRO-fit (p = .57, d = 0.33). All other interactions were non-significant (F < 1.30, p > .25).

Composite naming: By-participants and by-items analysis

GEE was conducted including all three factors (Context, Interview and System) as predictors, and Working Correlation Matrix set to 'exchangeable'. Using a Model-based estimator, all two- and three-way interactions were non-significant (p > .20) and these were removed from the model. In the new model [QIC = 464.55, $QICC$ = 464.74], all factors emerged as significant: (i) Context [$X^2(1)$ = 6.55, p = .010], revealing Detailed to be superior to Minimal [B = 0.64, $SE(B)$ = 0.25, $X^2(1)$ = 6.55, p = .010, $Exp(B)$ = 1.89, $95\%CI$ (1.16, 3.08)], (ii) Interview [$X^2(1)$ = 4.59, p = .032], with H-CI better-named than CI [B = 0.52, $SE(B)$ = 0.24, $X^2(1)$ = 4.59, p = .032, $Exp(B)$ = 1.68, $95\%CI$ (1.05, 2.71)], (iii) and System [$X^2(1)$ = 42.16, p < .001], as EvoFIT composites were more-accurately named than PRO-fit composites [B = -1.74, $SE(B)$ = 0.27, $X^2(1)$ = 42.16, p < .001, $Exp(B)$ = 0.18, $95\%CI$ (0.10, 0.30)].

Composite rating

Once again, the pattern of means for likeness ratings (Table 4) reflect those for correct naming. Repeated Measures ANOVA showed that there was a significant main effect of (i) CR [$F(1,17)$ = 15.52, p = .001, η_p^2 = .48], as composites produced following the Detailed CR technique were rated better than Minimal CR, (ii) Interview [$F(1,17)$ = 11.36, p = .004, η_p^2 = .40], as H-CI led to significantly better-rated composites than CI, and (iii) System [$F(1,17)$ = 12.83, p = .002, η_p^2 = .43], as composites were better rated for EvoFIT than PRO-fit. There was one significant interaction, System × Interview [$F(1,17) = 5.05$, $p = .038$, $\eta_p^2 = .23$]. All pairwise comparisons were significant ($p < .037$), but the rated likeness for EvoFIT was actually in the *opposite* direction to that of the main effect (i.e., CI was better than H-CI) following Detailed CR. Remaining interactions were non-significant ($F < 3.55$, $p > .07$).

Table 4. EvoFIT and PRO-fit mean composite rating (SD) by Context Reinstatement and Interview type

	Context Reinstatement†				
	Minimal		Detailed		
*System**	CI	H-CI	CI	H-CI	Mean
EvoFIT	1.7 (0.4)	2.0 (0.8)	2.2 (0.8)	2.1 (0.7)	2.0 (0.7)
PRO-fit	1.5 (0.4)	1.8 (0.5)	1.7 (0.5)	2.0 (0.7)	1.8 (0.5)
Mean	1.7 (0.6)		2.0 (0.7)		1.9 (0.7)

Note: † Significant main effect of CR, $p < .01$; *Significant main effect of System, $p < .01$; Significant System × Interview interaction, $p < .05$. The likeness rating scale was from 1 (poor likeness) to 7 (good likeness).

Discussion

Experiment 2 revealed an overall benefit for the Detailed CR manipulation which improved correct naming and likeness ratings of composites for both types of system. Also, composites produced via H-CI interview procedure were overall significantly better named and better rated than those following the CI technique. One exception to this was observed in the significant interaction in the rating data, suggesting that likenesses were perceived to be worse for H-CI (cf. CI) for EvoFITs under Detailed CR. Likeness ratings are a proxy to naming and so it is not surprising that differences between dependent variables (DVs) emerge; it is composite naming that is arguably more forensically relevant. EvoFIT composites were named and rated significantly better than PRO-fit composites.

These findings contrast with an unpublished study that employed the same design as Experiment 2, but did not find an advantage of Detailed over Minimal CR. Whilst means for naming and rating of PRO-fit composites were in the expected direction, albeit non-significant, likeness ratings for EvoFIT decreased for composites produced following Detailed (cf. Minimal) CR. This study, which we refer to as a Supplementary experiment (Fodarella, Marsh, Chu, Athwal-Kooner, & Frowd, 2019; see Appendix 1), involved an older, less-effective version of EvoFIT, which we thought may have led to more variable results. However, the advantage of CR for PRO-fit composites was weak, suggesting that another mechanism might be involved. One plausible mechanism is the degree of attention constructors paid to the environment. Glenberg's (1997) theory of environmental suppression suggests that environmental context is usually encoded automatically, except when individuals engage in additional conceptual processing. In Fodarella et al. (2019), poor encoding of the visual environment may stem from participants engaging in the process of the experiment, which itself did not directly refer to the environment, resulting in limited recall of contextual information and limited facilitation of face construction. This proposal is explored next.

In Experiment 3, attention of half the face constructors was explicitly directed to the environment during encoding. The approach is in line with past research indicating that an MCR benefit is more likely to occur when experimental instructions emphasise a so-called interactive encoding between study items (in this case, the target face) and the environmental context (see Hanczakowski, Zawadzka, & Coote, 2014; Hockley, 2008). Interactive encoding between items and context would ensure a stronger association between the two, which would presumably in turn ensure that they act as retrieval cues for one another during recall. While we are not directly manipulating the face to interact with the environment, the above evidence (Hanczakowski et al., 2014; Hockley, 2008) implies that increased attention to the environment should promote stronger context effects.

An attempt was also made in this experiment to facilitate the potential benefit of the environment on face construction in another way. The approach is based on the theory that memory for information can be facilitated by cued recall. As well as MCR, the extensive recall technique is an effective interviewing mnemonic of the Cognitive Interview, used to facilitate eyewitness recall (Fisher & Geiselman, 1992). Put simply, memory recall should be greater when multiple retrieval attempts are made rather than stopping after an initial memory search (Fisher & Geiselman, 1992). In Experiment 3, therefore, once participants had provided free recall of the environmental and internal context, they were asked to try to remember further information; specifically, participants were prompted (or 'cued') by open-ended questions based on the information they had recalled. For example, having mentioned the presence of chairs and tables in the room, participants would then be asked if they could recall further information about these items. It was hypothesised that this 'cued' technique would lead to a greater recall of the environment, which in turn should facilitate memory of the face as well as production of a composite.

The PRO-fit feature system was not used in this or in the following experiments, principally due to the low naming rates produced from its composites and the ensuing difficulty of then making sensible conclusions. The following experiments therefore focused on the EvoFIT system.

EXPERIMENT 3: Increasing the focus of attention on the environmental context

Stage 1: Composite construction

Design

The design was between-participants: 2 (Context Attention: Incidental, Intentional) × 3 (CR: Minimal, Detailed, Extensive). In the Intentional condition, participants' attention was directed to the environment by asking them to inspect the environment closely prior to viewing the target face; participants in the Incidental condition did not receive these instructions, with encoding of the environment carried out in the same (incidental) way as before. Prior to composite construction, CR was manipulated on three levels: (i) Minimal CR, (ii) Detailed CR and (iii) Extensive CR as (ii) with participants then invited to try to remember further information as part of a cued-recall format.

Participants

Sixty (20 males, 40 females; M_{age} = 30.4 years, SD_{age} = 9.4 years) UCLan staff and students volunteered. They were recruited opportunistically on the basis of being unfamiliar with the target faces.

Materials

Ten photographs of characters from the TV soap 'EastEnders' were target images (Ian Beale, Jane Beale, Jack Branning, Lauren Branning, Max Branning, Stacey Branning, Shirley Carter, Martin Fowler, Billy Mitchell and Jean Slater). They were of the same standard as in the previous experiments and were printed likewise.

Procedure

The construction procedure was the same as the three previous experiments except for the following differences. Prior to face encoding, participants were either asked to pay close attention to the café environment in which the face was to be subsequently shown (intentional encoding of context), or did not receive such an instruction (incidental encoding). If participants assigned to the former condition did not seem to study the room (which occurred about a third of the time, N = 11 / 30), the experimenter gave a prompt: "I will give you a little more time to look at the environment". The following day, participants were interviewed in one of three conditions prior to constructing the face via EvoFIT (incl. the instruction to select for the upper facial half in face arrays). Minimal and Detailed CR conditions were administered as before. Extensive CR followed the Detailed CR procedure, after which participants were asked questions about objects recalled in the environment. For instance, "You remembered tables and chairs. Can you say anything more about these?". The researcher prompted for further information about objects in the order in which they had been initially recalled.

Stage 2: Composite naming

Design, Materials and Procedure

The design was the same as in Stage 1. Materials were 60 composites and 10 target photographs, printed as before. The naming procedure was the same as in Experiments 1 and 2.

Participants

Sixty (15 males, 45 females; M_{age} = 41.6, SD_{age} = 12.7 years) staff and student volunteers were recruited opportunistically on the UCLan campus. Participants were recruited on the basis of being familiar with the target identities.

Stage 3: Composite rating

Design, Materials and Procedure

The Design was the same as in Stage 2, but within-participants. Materials were the same as in Stage 2, and Procedure the same as in the previous experiments.

Participants

Eighteen (9 males, 9 females; M_{age} = 26.7, SD_{age} = 5.8 years) UCLan staff and students volunteered. They were recruited on the basis of being unfamiliar with the target identities.

Results

Composite naming: By-participants analysis

Composites and targets were scored in the same way as in Experiments 1 and 2. Familiarity with the target identities was once again high, at 98.83%; mean composite naming is shown in Table 5.

Independent Samples ANOVA revealed a non-significant main effect of CR [$F(2,54) = 3.09$, $p = .05$, $\eta_p^2 = .10$], but a significant main effect of Attention [$F(1,54) = 11.91$, $p = .001$, $\eta_p^2 = .18$], with composites named significantly better when attention was directed to the environment (cf. Incidental). These two factors also interacted with each other [$F(2,54) = 3.34$, $p = .043$, $\eta_p^2 = .11$]. When attention was Incidental, there were no significant differences between CR conditions ($p >$ .89). However, when attention was Intentional, composites were named significantly better if constructed following Detailed [$t(18) = 3.14$, $p = .006$, $d = 1.40$] and Extensive [$t(18) = 2.98$, $p = .008$, $d = 1.33$] compared to Minimal CR; there was no significant difference between Detailed and Extensive CR [$t(18) = 0.09$, $p = .93$, $d =$ 0.05]. Also, when attention was Intentional (cf. Incidental), Detailed [$t(18) = 2.81$, $p = .012$, $d = 1.26$] and Extensive CR [$t(18) = 2.72$, $p = .014$, $d = 1.21$] produced significantly higher-named composites, but there was no reliable difference between the two Minimal CR conditions [$t(18) = 0.15$, $p = .88$, $d = 0.07$].

Table 5. Percentage of EvoFIT composites correctly named by Context Reinstatement and Attention

	Context Reinstatement			
*Attention**	Minimal	Detailed	Extensive	Mean
Incidental	23.0 (15.0)	23.0 (18.3)	22.0 (19.9)	22.7 (17.2)
Intentional	22.0 (14.8)	49.0 (22.8)	50.0 (25.8)	40.3 (24.7)
Mean	22.5 (14.5)	36.0 (24.1)	36.0 (26.6)	31.5 (22.9)

Note: * Significant main effect of Attention, $p < .01$; Significant Attention × CR interaction, $p < .05$. In parentheses are (by-participant) SD values.

Composite naming: By-participants and by-items analysis

The data was analysed using GEE, with Context and Attention as predictors. The Working Correlation Matrix was determined as 'exchangeable'. The best estimator emerged as model-based [*QIC* = 724.35, *QICC* = 715.69] and revealed Context to be a significant predictor of composite naming [$X^2(2) = 6.56$, $p = .038$]. Whilst both Detailed [$B = 0.69$, $SE(B) = 0.29$, $X^2(1) = 5.41$, $p = .020$, $Exp(B) = 1.99$, *95%CI* (1.11, 3.54)] and Extensive [$B = 0.69$, $SE(B) = 0.31$, $X^2(1) = 4.83$, $p = .028$, $Exp(B) = 1.99$, *95%CI* (1.08, 3.66)] were superior to Minimal, there was no significant difference between the two [$B = 0.00$, $SE(B) = 0.30$, $X^2(1) < 0.001$, $p = 1.00$, $Exp(B) = 1.00$, *95%CI* (0.55, 1.81)]. Attention also appeared to be a significant predictor [$X^2(1) = 10.89$, $p = .001$], with Intentional leading to better-named composites than Incidental [$B = 0.85$, $SE(B) = 0.25$, $X^2(1) = 11.40$, $p = .001$, $Exp(B) = 2.35$, *95%CI* (1.43, 3.85)].

In addition, the interaction between Context and Attention was significant [$X^2(2) = 7.14$, $p = .028$]. Further inspection of the data suggested that Minimal did not significantly differ whether attention was Incidental or Intentional [$B = -0.06$, $SE(B) = 0.36$, $X^2(1) = 0.03$, $p = .87$, $Exp(B) = 0.94$, *95%CI* (0.47, 1.92)]. When attention was Intentional, both Detailed [$B = -1.23$, $SE(B) = 0.41$, $X^2(1) = 8.91$, $p = .003$, $Exp(B) = 0.29$, *95%CI* (0.13, 0.66)] and Extensive [$B = -1.27$, $SE(B) = 0.40$, $X^2(1) = 9.85$, $p = .002$, $Exp(B) = 0.28$, *95%CI* (0.13, 0.62)] produced more-accurate naming than Minimal, whereas there were no significant differences between these conditions when attention was Incidental [Detailed vs Minimal: $B = 0.00$, $SE(B) = 0.40$, $X^2(1) < 0.001$, $p = 1.00$, $Exp(B) = 1.00$, *95%CI* (0.46, 2.19)]; Extensive vs Minimal: $B = -0.06$, $SE(B) = 0.43$, $X^2(1) = 0.02$, $p = .89$, $Exp(B) = 0.94$, *95%CI* (0.41, 2.19)].

Composite rating

Once again, the same basic pattern of results was observed in likeness ratings (Table 6). This time, Repeated Measures ANOVA revealed a significant main effect of CR [$F(2,34) = 64.34$, $p < .001$, $\eta_p^2 = .79$], with both Detailed [$t(17) = 12.42$, $p < .001$]

and Extensive [$t(17) = 9.70$, $p < .001$] giving rise to higher-rated composites than those for Minimal CR; as before, there was no reliable difference between Detailed and Extensive CR ($p = .11$). The main effect of Attention was significant [$F(1,17) = 164.18$, $p < .001$, $\eta_p^2 = .91$], with higher-rated composites for Intentional than Incidental.

Table 6. EvoFIT mean composite rating (SD) by Context Reinstatement and Attention type

	Context Reinstatement†			
*Attention**	Minimal	Detailed	Extensive	Mean
Incidental	1.3 (0.1)	1.4 (0.1)	1.3 (0.1)	1.3 (0.1)
Intentional	1.2 (0.1)	1.8 (0.2)	1.7 (0.2)	1.6 (0.3)
Mean	1.3 (0.1)	1.6 (0.3)	1.5 (0.2)	1.5 (0.3)

Note: † Significant main effect of CR, $p < .001$; * Significant main effect of Attention, $p < .001$; Significant Attention × CR interaction, $p < .001$. The likeness rating scale was from 1 (poor likeness) to 7 (good likeness).

The interaction between Attention and CR was also significant [$F(1,25) = 41.38$, $p < .001$, $\eta_p^2 = .71$]. Whilst composites constructed via Minimal CR were rated similarly whether attention was Incidental or Intentional [$t(17) = 0.68$, $p = .51$], those constructed via Detailed [$t(17) = 15.17$, $p < .001$] and Extensive CR [$t(17) = 7.36$, $p < .001$] were rated significantly higher following Intentional than Incidental encoding. For Incidental conditions, composites constructed via Detailed were rated significantly better than Minimal CR [$t(17) = 3.80$, $p = .001$], but there was no significant difference between Minimal and Extensive [$t(17) = 2.03$, $p = .06$] or between Detailed and Extensive CR [$t(17) = 0.41$, $p = .69$]. In contrast, composites produced following Intentional encoding were significantly better-rated following Detailed [$t(17) = 15.05$, $p < .001$] and Extensive CR [$t(17) = 9.76$, $p < .001$] compared to Minimal; Detailed and Extensive did not differ significantly [$t(17) = 1.97$, $p = .07$].

Discussion

When participants' attention was directed to the environment, composite naming and likeness ratings increased significantly compared to incidental encoding. However, this effect was driven by higher naming and ratings of composites created under Detailed and Extensive CR. So, the advantage of Detailed CR (cf. Minimal CR) emerged when the environment had been encoded *intentionally*, as reflected in both naming and likeness ratings. When attention to the environment was incidental, Detailed and Extensive CR were not effective in increasing composite naming, although the former condition reliably increased (with a small *MD*) for likeness ratings (cf. Minimal). Again, this result indicates the influence of seeing (but not attempting to deliberately remember) the environment, an effect that was also observed in likeness ratings in Experiments 1 and 2. Overall, these results imply that, in two out of three experiments, participants had encoded the environment to some extent despite not having been specifically asked to do so, and this influence positively affected ensuing likenesses.

The composite naming and likeness ratings also revealed that Minimal CR did not differ reliably between incidental and intentional attention. This finding would appear to be sensible since Minimal CR following intentional encoding did not then make active use of the environment. This suggests that trying to remember the environment at the point of encoding is not sufficient in itself to facilitate an ensuing composite (as assessed by naming or likeness): what is necessary is to *recall* the environment. A final noteworthy result relates to Extensive CR. It was predicted that this condition would promote more effective composites than Detailed CR. This was not the case for composite naming and likeness ratings in either incidental or intentional encoding. Thus, attempts to recall more information about the environment, even if there should have been more information available to recall (esp. following intentional encoding), does not seem to have influenced the ensuing composites. The implications of these findings are discussed in greater depth in the General Discussion.

In the experiments conducted so far, results are not consistent when participants' attention was not directed to the environment: there was advantage of Detailed CR (cf. Minimal CR) in Experiments 1 and 2, as assessed by composite naming, and Experiments 1, 2 and 3 by likeness ratings. In the next section, we assess the overall magnitude and reliability of Detailed CR by combining data across experiments using meta-analysis. Our main interest is in the forensically relevant measure, naming, and so this aggregated type of analysis focused on this DV. In this section, we focus not only on correct naming but also consider incorrect naming of composites. In addition, to provide evidence of whether Detailed CR is effective by improving memory of the face, we assess the quantity of information recalled about the target face by constructors in the previous experiments.

META-ANALYSES AND ADDITIONAL ASSESSMENTS

Method

Studies and Procedure

Eleven comparisons between Minimal versus Detailed CR were available for meta-analyses, derived from Experiments 1 to 3 and Supplementary experiment (Fodarella et al., 2019), for correct naming, incorrect naming and face recall. Comparisons include only incidental encoding of the environment (i.e., cases where attention had been directed to the café in Experiment 3 were omitted). Cohen's *d* was used as the measure of effect size in each comparison along with its standard error (SE).

For an analysis of face recall, participants' free face recall elicited prior to composite construction was coded by assigning a value of 1 for each unit of information (UOI) recalled. For example, "<u>small</u>, <u>brown</u> eyes" would be counted as two UOI, and "eyebrows were <u>far apart</u>, <u>low</u> and <u>quite straight</u>" as three. Information regarding details other than the face (e.g., clothing, jewellery and shoulders) were excluded, whilst subjective information about the face, such as "<u>pleasant</u>, <u>good-looking</u> face" (two UOI) or "<u>quite a friendly</u> face" (one UOI), were included; information relating to either of these categories was recalled by around 1 in 4 participants. Using this scoring procedure, total face recall was calculated for each facial composite across experimental conditions. To ensure coding consistency, the same two experimenters coded recall. Both coders were blind to the experimental conditions under which composites had been constructed and participant recall was presented in a random order (for both experimental conditions and target identity). After coding, scores were compared and differences resolved by discussion; this occurred only on two occasions, and thus inter-rater reliability emerged at 100%.

Four main meta-analyses were conducted in Microsoft Excel using a template provided by Neyeloff, Fuchs, and Moreira (2012). The meta-analyses were for (i) correct naming by-participants, (ii) correct naming by-items, (iii) incorrect naming and (iv) face recall by-participants. Initial analysis of the first three sets of data included all eleven comparisons, while the remaining two analyses explored the effect by composite system. For face recall data, the initial analysis included the seven comparisons in which Detailed CR reliably increased correct naming. The second analysis included data from Experiment 3, and the third analysis also included data from Fodarella et al. (2019). Mainly due to low statistical power, results of face recall are presented overall (cf. by composite system). It was expected that this information would only increase under Detailed CR when the meta-analysis included experiments in which the manipulation had been successful—that is, when correct naming had increased significantly. The effect was likely to decrease when including additional data (i.e., in the second and third meta-analysis).

Results

Correct naming

Table 7 lists comparisons involved in the first meta-analyses: overall correct naming scores (by-participants) between Detailed and Minimal CR. For each comparison, mean values are presented with the associated Cohen's *d* standardised measure of effect size. The Median is also presented as an alternative measure of central tendency. As can be seen, each of these percentage-correct values is very similar to the mean, with fairly low inter-quartile range (IQR), indicating that analyses were not unduly influenced by outliers.

Table 7. Summary of comparisons included in the meta-analyses: correct naming (by-participants) between Detailed and Minimal CR

	Detailed CR			*Minimal CR*			
Study	Mean	Median	IQR	Mean	Median	IQR	Cohen's d
Experiment 1							
EvoFIT	33.8	40.0	12.5	23.8	25.0	17.5	0.7
PRO-fit	13.8	20.0	12.5	6.3	10.0	10.0	1.0
Supplementary Experiment							
EvoFIT (CI)	20.3	25.0	18.8	23.4	31.3	28.1	-0.2
EvoFIT (H-CI)	29.7	31.3	15.6	34.4	37.5	18.8	-0.3
PRO-fit (CI)	6.3	6.3	12.5	4.7	0.0	12.5	0.2
PRO-fit (H-CI)	9.4	12.5	12.5	7.8	12.5	12.5	0.2
Experiment 2							
EvoFIT (CI)	26.3	25.0	12.5	10.0	12.5	12.5	1.2
EvoFIT (H-CI)	31.3	37.5	21.9	22.5	25.0	28.1	0.5
PRO-fit (CI)	5.0	0.0	12.5	2.5	0.0	0.0	0.4
PRO-fit (H-CI)	6.3	0.0	12.5	6.3	6.3	12.5	0.0
Experiment 3							
EvoFIT	23.0	25.0	32.5	23.0	30.0	17.5	0.0

When both systems were included in the meta-analysis, the Random-effects model indicated that Detailed CR reliably increased correct naming over Minimal CR [$Q(10) = 9.98$, $I^2 = -0.20$, $SE(d) = 0.14$, 95%$CI(d)$ (0.05, 0.61)] with a small effect size (ES) ($d = 0.33$). When analysed by composite system, the benefit of Detailed CR over Minimal CR only emerged significantly for PRO-fit [$Q(4) = 3.84$, $I^2 = -4.26$, $SE(d) = 0.16$, 95%$CI(d)$ (0.01, 0.64)], again with a small ES ($d = 0.33$); the result for EvoFIT failed to reach significance [$Q(5) = 5.04$, $I^2 = 0.85$, $SE(d) = 0.24$, 95%$CI(d)$ (-0.14, 0.78)], again the ES was small ($d = 0.32$). To increase statistical power, the latter comparison was analysed using a one-tailed Independent Samples t-test that compared all participant data for EvoFIT composites: Detailed CR ($M = 28.24$, $SD = 15.01$) was now significantly higher than Minimal CR ($M = 22.36$, $SD = 16.81$) [$t(106) = 1.92$, $p = .029$, $d = 0.27$].

A meta-analysis by-items, using naming scores calculated for each composite (cf. each participant), was reliable for these 11 comparisons [$Q(10) = 10.53$, $I^2 = 5.02$, $SE(d) = 0.10$, 95%$CI(d)$ (0.11, 0.52)], again with a small ES ($d = 0.32$). For completeness, separate analyses were conducted for each system, but neither reached significance [EvoFIT: $Q(5) = 4.52$, $I^2 = -10.73$, $d = 0.30$, $SE(d) = 0.19$, 95%$CI(d)$ (-0.07, 0.67); PRO-fit: $Q(4) = -0.13$, $I^2 = 3290.21$, $d = 0.16$, $SE(d) = 0.09$, 95%$CI(d)$ (-0.02, 0.34)]. This analysis using smaller subsets was not entirely unexpected since by-items analyses tend to be statistically weaker in composite studies due to there usually being more participants than items in an experiment, meaning that SE estimates emerge smaller. The overall result using all data indicates that the Detailed CR advantage generalises to other items.

Incorrect naming

A Random-effects model showed that there was no significant difference in incorrect naming for EvoFIT composites [$Q(5) = 4.77$, $I^2 = -4.83$, $SE(d) = 0.13$, $95\%CI(d)$ (-0.37, 0.14)], with a small ES ($d = -0.11$). The negative ES indicates that there was a reduction in incorrect names, albeit non-significant. For PRO-fit, the trend was in the opposite direction, that is, an increase in incorrect names, but again this was non-significant [$Q(4) = 3.96$, $I^2 = -0.96$, $SE(d) = 0.20$, $95\%CI(d)$ (-0.26, 0.54)], with a small ES ($d = 0.14$). As the two systems have an effect in the opposite direction, when combined, the overall effect is even smaller ($d = -0.01$) and again non-significant [$Q(10) = 9.77$, $I^2 = -2.34$, $SE(d) = 0.12$, $95\%CI(d)$ (-0.23, 0.22)].

Face recall

Mean (and SD) for face recall for each comparison across experiments was calculated along with Cohen's *d* (Table 8). A Random-effects model involving the six comparisons under which correct naming reliably improved for Detailed CR (cf. Minimal) indicated a significant increase in face recall [$Q(5) = 5.52$, $I^2 = 9.40$, $SE(d) = 0.14$, $95\%CI(d)$ (0.22, 0.77)], with a medium ES ($d = 0.49$). Including data from Experiment 3, when no advantage was observed by correct naming, the effect is now medium in size ($d = 0.36$) and the meta-analysis was marginally significant [$Q(6) = 5.97$, $I^2 = -0.42$, $SE(d) = 0.19$, $95\%CI(d)$ (-0.01, 0.73)]. When also including data from the Supplementary experiment, ES decreases even further ($d = 0.16$) and the analysis is no longer even marginally significant [$Q(10) = 9.99$, $I^2 = -0.05$, $SE(d) = 0.15$, $95\%CI(d)$ (-0.14, 0.46)].

Table 8. Mean (and SD) face recall and Cohen's *d* by CR across experiments

	Context Reinstatement		
Study	*Detailed CR*	*Minimal CR*	Cohen's d
Experiment 1			
EvoFIT	9.4 (2.6)	7.4 (1.8)	0.9
PRO-fit	11.3 (4.7)	11.1 (3.5)	0.0
Supplementary Experiment			
EvoFIT (CI)	9.0 (4.1)	10.5 (1.9)	-0.5
EvoFIT (H-CI)	11.6 (4.9)	13.6 (4.9)	-0.4
PRO-fit (CI)	11.3 (3.9)	10.5 (2.7)	0.2
PRO-fit (H-CI)	12.9 (5.3)	13.6 (2.6)	-0.2
Experiment 2			
EvoFIT (CI)	13.4 (3.0)	11.9 (4.7)	0.4
EvoFIT (H-CI)	12.8 (5.1)	10.8 (4.1)	0.4
PRO-fit (CI)	12.3 (2.7)	10.9 (3.0)	0.5
PRO-fit (H-CI)	13.3 (2.7)	10.1 (3.8)	1.0
Experiment 3			
EvoFIT	9.5 (3.1)	10.7 (2.1)	-0.5

Discussion

Meta-analyses of Experiments 1 to 3 and the Supplementary experiment (Fodarella et al., 2019) revealed that, for incidental conditions, there was a reliable benefit of Detailed over Minimal CR for all participants with a small effect size. An effect size of similar magnitude emerged separately for PRO-fit and for EvoFIT composites. This advantage was reliable for PRO-fit and, with a t-test involving all participant data, for EvoFIT. The by-items analyses for correct naming supported this reliable finding with all data combined but, presumably due to limited power, not for systems considered separately. Incorrect naming did not significantly differ between conditions, neither for all data combined nor for each system, a finding indicating that Detailed CR does not increase misidentification of composites. The face recall analyses showed that only in those cases when Detailed CR was successful in increasing composite identifiability (cf. Minimal CR) was an overall effect found for face recall. When including the 'unsuccessful' comparisons, the overall effect of face recall became weaker (incl. data from Experiment 3) and weaker again (incl. data from Fodarella et al., 2019), thereby supporting the idea that the CR effect, when successful, is driven by an increase in memory recall of the target face.

We argued earlier that an absent or weak advantage for Detailed CR might stem from insufficient encoding of the encoding environment. In the next experiment, participants again encoded the environment incidentally (as was the design of Experiments 1 and 2, and for components of Experiment 3) but this time we attempted to improve memory for the environment more-naturally. Previously, participants followed the experimenter into the room in which the target face had been seen. However, this procedure may not promote good encoding of the environment, and so participants here were invited to enter the room in front of the researcher. In doing so, participants should naturally take in details of the relevant context in order to navigate to the table indicated by the experimenter. This should also more-closely reflect how eyewitnesses encode an environment in the real world (as opposed to the intentional encoding that we used in the previous experiment).

We were also interested in assessing a simpler version of the memory technique, one that does not require participants to recall their emotional state at the time of encoding. Crimes for which composites are usually constructed involve considerable stress for an observer (esp. in cases of assault). Recalling such evocative information may be particularly traumatic, potentially leading to anxiety and the subsequent effect of inhibiting facial composite construction (Davies, 2009); conversely, attempts to reduce anxiety also seem to promote more-identifiable composites (Martin, Hancock, & Frowd, 2017). We reasoned that recall of the environment on its own should be sufficient to facilitate face construction (i.e., without the need for recall of psychological/emotional states).

We also thought it could be beneficial to law enforcement to assess the effectiveness of CR for a novel method of face construction. Recent technological advances have allowed witnesses to construct a composite themselves in their own time (e.g., Martin et al., 2018). This development allows composites to be used in investigations of less serious crime (e.g., minor theft, vandalism or anti-social behaviour), cases where police practitioners may not have the time (e.g., Alison, Doran, Long, Power, & Humphrey, 2013) to interview witnesses to create a composite. As such, the production team behind EvoFIT created a 'self-administered' version (www.EvoFIT.co.uk). This novel approach was designed to be functionally equivalent to the 'face-to-face' method used by police practitioners (as followed in the current experiments), with witnesses taken through the same procedure by following written instructions on-screen in their own home.

In light of this development, the following experiment assessed whether the advantage of Detailed CR would be apparent for participants who constructed the face in the normal manner via face-to-face interview, or by themselves. The experiment recruited residents from a small town in the UK (cf. university staff and students) to extend findings to other participant pools, with target encoding taking place in a small office (unfamiliar to participants) and face construction (for both face-to-face and self-administered) elsewhere in a relaxed home environment. We were principally interested in identification as a measure of success (cf. ratings of likeness), and so the resulting composites were assessed by naming only.

EXPERIMENT 4: Facilitating more-naturalistic encoding of the environment

Stage 1: Composite construction

Design

The design was between-subjects with two experimental factors: 2 (CR: Minimal, Detailed) × 2 (Face construction: Face-to-face, Self-administered). It was the same as the previous experiments by CR (Minimal vs. Detailed), except that Detailed CR did not involve recall of the participant's emotional state at encoding, and half of the composites were constructed using a self-administered procedure (with the other half produced with the assistance of the experimenter, as before). The other change relates to the room used for target encoding. It was a small office, previously unseen by the recruited participants, with potentially useful recall cues: computer, chair, desk, bookcase, stationery, etc. To facilitate encoding of the environment, participants were invited to enter the room in front of the experimenter and navigate to where they were instructed to sit.

Participants

Thirty-two (10 males, 22 females; M_{age} = 24.0 years, SD_{age} = 7.8 years) local residents of a small town in the UK (Whitchurch, Shropshire) volunteered. They were recruited opportunistically, on the basis that they were not familiar with the target faces.

Materials

Eight photographs of current EastEnders characters were target images (Ian Beale, Jack Branning, Lauren Branning, Max Branning, Stacey Branning, Shirley Carter, Billy Mitchell and Jean Slater), sourced and printed as before to the same standard.

Procedure

Face construction was the same as Experiments 1 to 3, except for the following differences. As before, participants were randomly assigned to CR (Minimal, Detailed), and now to Face construction (Face-to-face, Self-administered). Face-to-face construction proceeded as described previously. Those assigned to self-administered construction wrote down on a piece of plain paper what they could remember of the environmental context and then, on the reverse side, provided a free-recall description of the previously-seen face. When complete, participants followed the on-screen instructions on a laptop computer as described above to construct an EvoFIT composite.

Stage 2: Composite naming

Design, Materials and Procedure

The two-factor design was the same as in Stage 1. Thirty-two composites and eight target photographs were printed in greyscale (8 cm × 10 cm). Except for variation in the two experimental factors (CR and Face construction), the naming procedure was the same as that described previously.

Participants

Thirty-two (13 males, 19 females; M_{age} = 20.0, SD_{age} = 1.5 years) student volunteers were recruited opportunistically on the UCLan campus. Participants were recruited to be familiar with the target identities.

Results

Responses to composites and target photographs were scored for accuracy. All participants correctly named all eight target photographs, except for one participant who named seven, and so familiarity with the target set was high, at 99.7%. Table 9 provides a summary of composite naming, suggesting benefit for the Detailed CR procedure.

Table 9. Percentage of EvoFIT composites correctly named by Context Reinstatement and by Face construction

	Context Reinstatement†		
*Face construction**	Minimal	Detailed	Mean
Face-to-face	54.7 (9.3)	71.9 (12.9)	63.3 (14.0)
Self-administered	42.2 (9.3)	50.0 (11.6)	46.1 (10.9)
Mean	48.4 (11.1)	60.9 (16.4)	54.7 (15.1)

Note: † Significant main effect of CR, $p < .005$; * Significant main effect of Face construction, $p < .001$. In parentheses are (by-participant) SD values.

Independent Samples ANOVA revealed a significant main effect of CR [$F(1,28) = 10.54$, $p = .003$, $\eta_p^2 = .27$], with composites named higher when constructed following Detailed than Minimal CR ($d = 0.92$). Face construction was also significant [$F(1,28) = 20.00$, $p < .001$, $\eta_p^2 = .42$], with composites constructed face-to-face named higher than when self-administered. The interaction between CR and Face construction was not reliable [$F(1,28) = 1.48$, $p = .23$, $\eta_p^2 = .05$].

Generalised Estimating Equations (GEE)

GEE was used to analyse the data, and both Context and Construction were included as predictors. The Working Correlation Matrix was set to 'exchangeable' and the estimator was model-based. The interaction between the factors was not (p = .17) and was removed from the model; subsequently both factors emerged significant in the new model [QIC = 343.06, $QICC$ = 346.80]: (i) Context [$X^2(1)$ = 10.82, p = .001], with Detailed superior to Minimal [B = -0.52, $SE(B)$ = 0.16, $X^2(1)$ = 10.82, p = .001, $Exp(B)$ = 0.59, $95\%CI$ (0.43, 0.81)], and (ii) Construction [$X^2(1)$ = 20.18, p < .001], with the Face-to-face superior to Self-administered method [B = -0.71, $SE(B)$ = 0.16, $X^2(1)$ = 20.18, p < .001, $Exp(B)$ = 2.04, $95\%CI$ (1.49, 2.78)].

Discussion

Experiment 4 revealed that Detailed CR was still effective (cf. Minimal CR) when participants were asked to recall only the environmental context (i.e., they omitted recall of their psychological state at encoding), thus indicating a persistent advantage of Detailed CR with recall of physical environmental cues only. Although overall results indicate that self-administered composites were less identifiable than those produced face-to-face with the interviewer (experimenter), the positive effect of Detailed CR was consistent for both methods of face production. Thus, our effort to ensure (as far as possible) that participants encoded the environment sufficiently seems to have been successful.

Our working hypothesis has been that detailed recall of the environment allows constructors to achieve a better *memory* of the target face, the knock-on effect of which is for them to achieve a more accurate composite of this identity. Indeed, in the above meta-analysis, constructors' total recall of the face was used as an index of memory performance, and it was shown that increased recall under Detailed (cf. Minimal) CR related positively to a reliable increase in correct naming of composites. However, an alternative (and not necessarily mutually-exclusive) explanation is that Detailed CR facilitates face recognition, such as observed when context is reinstated using cues associated with the target (e.g., Shapiro & Penrod, 1986). A simple test of this theory would be to reverse the order of interviewing mnemonics: with free recall of the face carried out first followed by Detailed CR. If Detailed CR improves face recognition, then correct naming of composites would be expected to increase (cf. Minimal CR); yet, if the mechanism is mediated mainly by improvement of face recall, then no such benefit should be observed. We investigated this proposal in the current experiment using the same design as Experiment 4 but with the order of these two mnemonics reversed.

EXPERIMENT 5: Timing of detailed recall of the environmental context

Stage 1: Composite construction

Design

The design was between-subjects with two experimental factors: 2 (CR: Minimal, Detailed) × 2 (Face construction: Face-to-face, Self-administered). For Detailed CR, participants freely recalled the target face first and then the environmental context; for other participants, only free recall of the face was requested. As Experiment 4, participants entered the unfamiliar room used for encoding prior to the experimenter, to provide suitable opportunity for encoding of context; also, for Detailed CR, participants again recalled the physical, environmental (but not psychological) context.

Participants

Thirty-two (14 males, 18 females; M_{age} = 24.9 years, SD_{age} = 9.5 years) participants volunteered. Participants were recruited via volunteer sampling on the basis that they were unfamiliar with the target faces.

Materials

Targets were eight photographs of characters from the ITV Coronation Street soap (Peter Barlow, Carla Connor, Tyrone Dobbs, Tracey McDonald, David Platt, Kirk Sutherland, Leanne Tilsley and Sally Webster). Photographs were of the same standard as in the previous experiments and were printed likewise.

Procedure

Face construction was the same as in Experiment 4, except for the following differences. To keep the construction procedure closely aligned with the first three experiments, participants (randomly) assigned to self-administered construction were asked to verbally describe the target face as well as the environmental context, with the experimenter writing down recall (cf., Experiment 4, with participants writing down this information). The encoding environment was a Multi-Faith Centre on the UCLan campus, for approximately half of the participants, and a café in a local town, both rooms were rich in environmental cues and unfamiliar to participants. Face construction requested all participants to "think back" to when the face had been seen the previous day. For Detailed CR, the order of instructions was reversed: here, free recall of face was requested first followed by free recall of environment; Minimal CR, as before, only involved free recall of the face.

Stage 2: Composite naming

Design, Materials and Procedure

The two factorial design was the same as in Stage 1, and materials were 32 composites and eight target photographs, printed as before. The procedure was also the same as before, composite naming and then target naming, but was extended in one way. It was anticipated that if the benefit of Detailed CR was driven in part by improvement in face recall, the observed effect would be weaker. As such, statistical power was increased by asking participants to name their assigned set of composites for a second time, after having seen the target photographs; this is another commonly used procedure in composite research (e.g., Frowd, Bruce, Ross, McIntyre, & Hancock, 2007). For this second presentation of composites, which we refer to as 'cued' naming (cf. 'spontaneous' naming, first presentation), images were presented again in the same (random) order.

Participants

Thirty-six (14 males and 22 females; M_{age} = 28.1 years, SD_{age} = 13.3 years) participants were recruited via opportunity sampling on the UCLan campus on the basis of being familiar with the targets.

Results

All participants correctly named all eight target photographs, and so familiarity with the target set was at 100%. Descriptive statistics (Table 10) revealed that, for the initial Spontaneous naming task, correct naming scores reduced slightly under Detailed (cf. Minimal) CR for face-to-face construction, but the reverse was observed for self-administered construction. There was a similar outcome by CR, albeit with a higher level of naming, for Cued naming.

MANOVA revealed a significant main effect of Face construction [$F(2, 31) = 7.88$, $p = .002$, $\eta_p^2 = .34$], as composites were correctly named higher using face-to-face (M = 66.3%, SD = 12.0) than self-administered construction (M = 51.4%, SD = 19.4); here, Pillai's Trace is reported, based on Levene's Test of Equality of Error Variances ($p < .05$). There was no significant effect of either CR [$F(2, 31) = 0.61$, $p = .55$, $\eta_p^2 = .04$] or the interaction between CR and Face construction [$F(2, 31) = 0.39$, $p = .68$, $\eta_p^2 = .02$].

Table 10. Percentage of EvoFITs correctly named by Context Reinstatement, Face Construction and Method of composite naming

	Context Reinstatement	
	Minimal	Detailed
Spontaneous Naming		
*Face construction**		
Face-to-face	43.1 (14.1)	37.5 (8.8)
Self-administered	26.4 (9.8)	33.3 (24.2)
Cued Naming		
*Face construction**		
Face-to-face	94.4 (11.0)	90.3 (13.7)
Self-administered	72.2 (16.3)	66.7 (25.8)

Note: * Significant main effect of Face construction, $p < .005$. In parentheses are (by-participant) SD values.

Generalised Estimating Equations (GEE)

The predictors Context and Construction were analysed using GEE, with accuracy of spontaneous composite naming as DV. Model-based estimator was used and the Working Correlation Matrix determined to be 'exchangeable'. The model was not significant [QIC = 380.17, $QICC$ = 381.98]; both the two factors (p > .80) and the interaction between the two (p > .41) emerged as non-significant.

Discussion

Experiment 5 revealed that Detailed CR was not effective at increasing correct naming of composites when used after participants had described the target face, and that this null effect emerged irrespective of whether the face was constructed using the conventional face-to-face procedure, or when self-administered. The result suggests that correct composite naming improves for Detailed (cf. Minimal) CR due to increases in face recall rather than face recognition. Mirroring Experiment 4, it was also found that composites were constructed more effectively when participants worked with the experimenter than alone.

General Discussion

The current work involved five experiments that examined environmental-context techniques as retrieval cues to potentially improve a person's ability to create a facial composite from memory. Asking constructors to recall both the visual environment and their psychological context where a target face had been seen was successful in increasing the effectiveness of facial composites produced from a typical recognition system (EvoFIT) and a typical feature system (PRO-fit). Environmental context, however, only seemed to be valuable in aiding memory if participants ('witnesses') paid sufficient attention to it at encoding. Experiment 2 showed that CR procedures could be effective in combination with a further interviewing technique, H-CI. The work also revealed that employing an extensive recall technique, one that encouraged additional retrieval of the environment using cued-recall questions, was not successful in rendering composites more identifiable.

The first two experiments used a holistic system (EvoFIT) and a feature system (PRO-fit). The main reason for including both types of systems was to investigate whether CR interviewing techniques could be successful for both types, rather than comparing the systems *per se*. Other research has focused on the effectiveness of systems (e.g., see Frowd et al., 2015 for meta-analysis), and here we find the same overall outcome, that recognition systems are more effective than feature systems.

Despite naming rates generally having increased over time with these newer recognition systems (for a review, see Frowd, 2017), composites still contain error; a natural result since they are constructed from memory. Therefore, for a composite to be most effective in aiding police to catch criminals, it is important to develop new techniques that help to facilitate memory, such as making use of contextual cues. To our knowledge, limited research (i.e., Davies & Milne, 1985, and Ness & Bruce, 2006) has applied CR techniques for the purpose of enhancing composites, and our results from Experiment 1 mirror, as well as extend, previous findings involving the forensic use of contemporary composite systems. Compared to Minimal CR—which consisted of an instruction to think back to the time of encoding and to freely recall the face—both recognition and feature composites were rated as significantly better likenesses following Physical CR, but achieved overall best likeness ratings and identification (correct naming) rates following Detailed CR[5]. Our findings are in line with the Encoding Specificity theory (Tulving & Thompson, 1973): as participants were first involved in recalling the environmental and internal con-

[5] Detailed CR is equivalent to Mental CR from Davies and Milne's (1985) experiment (see Footnote [1]).

text, it seems as though this information acted as associative retrieval cues, facilitating access to facial memory. Physical CR did not have a facilitative effect. The interviewing procedure prior to composite construction did not differ between Minimal and Physical conditions; that is, participants only actively recalled the *target face*, but did not engage in intensive memory recall of the environment (cf. Detailed CR). Therefore, the environmental context as a physical retrieval cue in itself does not appear to be strong enough to facilitate memory. Since Physical CR poses practical and ethical problems for policing, although there are potential ways to overcome this issue (see Ness & Bruce, 2006), Detailed CR is a more convenient procedure to implement, and may be less traumatic for victims. Davies and Milne's (1985) results also indicate that Physical CR was not as effective as Mental (Detailed) CR. Therefore, the following experiments focused on Detailed CR.

Experiment 2 also found that Detailed CR was successful in increasing correct naming of composites, overall. However, this finding was not supported by a Supplementary experiment (Fodarella et al., 2019), one involving a similar design, albeit using an archaic version of EvoFIT, with Detailed CR found not to be as effective. It would appear that inconsistencies may have emerged in the visual encoding of the environment. In line with the theory of environmental suppression (Glenberg, 1997), participants may have been preoccupied with the procedure of the experiment, which itself did not direct participants' attention to the café in which the target face had been seen, the result of which seems to be insufficient encoding of the environmental context.

In Experiment 3, attention was specifically directed to the environment for intentional encoding, or not, giving rise to the previous type of encoding, incidental. Findings were sensible: Detailed CR was only effective in increasing composite naming following intentional encoding. As in Fodarella et al. (2019), Detailed CR was not effective in incidental conditions, indicating that inconsistencies are due to a general lack of attention to the environment. Given that directed (intentional) encoding of the environment seems unnatural, Experiment 4 attempted to improve memory of the environment more realistically, by inviting witnesses to enter the testing room first. Results confirmed that Detailed CR reliably improved composite naming (cf. Minimal), for two methods of face production, implying that participants had, without prompting, encoded the environment to a greater extent, a situation that is closer to real life. The advantage for Detailed CR in Experiment 4 was large in size (d = 1.08), although it was not as large as that found following intentional encoding in Experiment 3 (d = 1.40), as one might expect. Since Detailed CR was also effective in Experiments 1 and 2, it seems that participants there paid sufficient attention to the environment without having been specifically asked to do so, while in Experiments 3 and Fodarella et al. (2019), this appears not to have been the case.

So, the nature of environmental encoding governs the stability of the effect. Meta-analyses of all conditions in which the environment had been encoded incidentally indicate that recalled context cues reliably improved correct naming of composites for both systems, with a small effect size (d = 0.25). The analyses also indicated that the overall effect is generalisable not only to other participants but also other target faces (items). In a real-life crime, it is thought that a witness or victim is unlikely to intentionally encode the environment around them, and so it is encouraging to observe that incidental encoding has subsequent benefit to face construction. Perhaps more importantly, even when Detailed CR was not effective, it did not reliably reduce correct naming of composites, nor alter incorrect naming, so there is no apparent disadvantage in using this technique in the real world.

Experiment 4 involved a more naturalistic encoding of the environment, by requesting participants enter the testing room in front of the experimenter, and sit in a specific chair across the room, thus encouraging greater engagement in and thus encoding of the environment. This experiment also emulated the more realistic scenario where a constructor's psychological state would not be recalled (cf. Experiments 1-3): recall of the visual environment alone was found to be sufficient to facilitate face construction. This has important practical implications since witnesses and victims (esp. of serious crimes such as assault) would find having to recall their psychological state traumatic, which in turn is likely to inhibit face construction (Davies, 2009). This version of Detailed CR (i.e., without reference to psychological state) was advantageous for the conventional face-to-face method to construct the face, as well as for a novel 'self-administered' procedure in which constructors engaged in environmental recall, face recall and face construction on their own. We do acknowledge that we have not considered the effect of recalling the constructor's psychological state alone, and it would seem sensible from a theoretical perspective to consider this possibility in future work (although based on the above result from Detailed CR, it is unlikely that recall of internal state would be effective).

The underlying mechanism that drives Detailed CR to increase composite naming seems to be an increase in memory recall of facial features of the target. Meta-analyses of the face recall data revealed that, for those experiments in which Detailed CR was successful in improving correct naming, there was a reliable increase in the amount of information recalled. When conditions were included in which Detailed CR was not successful—specifically, those conditions when participants supposedly had not paid sufficient attention to the environment (Experiment 3 incidental conditions and Fodarella et al., 2019)—the reliable increase in face recall diminished. Experiment 5 provides further evidence: when recall of environment and face were reversed, Detailed CR was not effective in facilitating face construction. If Detailed CR turned out to facilitate face recognition, one would expect the procedure to remain effective, such as in cases when cues associated with the target facilitated face recognition (Shapiro & Penrod, 1986). However, as this particular order of recall was unsuccessful, it seems likely that the underlying mechanism of Detailed CR is mediated by an increase in face recall rather than face recognition.

Experiment 2 included an additional interviewing technique to potentially increase composite effectiveness further: an H-CI, upon which participant-witnesses reflected and then made additional personality-type judgments for their target face. The H-CI significantly increased composite effectiveness for both systems (cf. CI), replicating past research (e.g., Frowd et al., 2008, 2015). Whilst the more standard CI face-recall procedures promote face description through feature recall and are likely to interfere with recognition (Dodson, Johnson, & Schooler, 1997; Frowd & Fields, 2011), the H-CI is believed to promote holistic face recall and in turn composite effectiveness. In Experiment 2, H-CI was effective in conjunction with Detailed CR, leading to the most effective composites compared to other conditions. Both of these procedures may be effective in combination by guiding witness attention to different aspects of the face: the H-CI shifts the focus towards the central part of the face, whilst Detailed CR provides better memory for facial features. If memory is improved for both the whole face and individual features then it would seem reasonable that the two techniques combined could improve face construction. It is now established that shifting a witness's focus of attention to the central part of the face aids composite construction (e.g., Frowd et al., 2008). On the other hand, improved memory for facial features is likely to aid discrimination between features—such as, eyes and brows. As the region around the eyes plays a central role for familiar face recognition (O'Donnell & Bruce, 2001), relevant here to composite naming, it seems plausible that better memory for this area should assist in constructing a better likeness, which in turn would increase subsequent identification (Ellis, Shepherd, & Davies, 1980).

In Experiment 3, Detailed CR was considered along with another interviewing mnemonic to potentially facilitate memory further, extensive retrieval (Fisher & Geiselman, 1992), in this case extensive recall of the environment. Rather than stopping after one, initial memory search, it is thought that multiple retrieval attempts elicit greater overall recall. Using Detailed CR, participants were invited to freely recall the environmental and internal context. In Experiment 3, extensive retrieval was encouraged using open-ended questions about the encoding environment. These data, however, suggest that further recall was no more effective for improving composite effectiveness (by naming or likeness ratings) compared to Detailed CR.

It is not entirely clear why this additional mnemonic was not more effective. Constructors in this condition each recalled further detail when probed, showing that the technique elicited more information compared to the initial, free recall. It is possible that this additional information simply did not facilitate facial memory any further, maybe because the information was less relevant than free recall, or perhaps participants had already visualised this information during free recall but did not verbalise it to the experimenter despite being asked to recall as much information as possible. Our findings are somewhat in line with results from Campos and Alonso-Quecuty (1999) who found that four multiple retrieval attempts ("try again") were not as effective in increasing correct recall as other retrieval strategies—CR, change perspective, recall in a different order, and other techniques. Although participants in our Experiment 3 were not specifically asked to "try again" (as in Campos & Alonso-Quecuty, 1999), our cued-based questions may not be as effective as other retrieval mnemonic strategies. If this is the case, one straightforward way to trigger new information may be to request recall in a different spatial order, perhaps starting from the location of last item remembered. Future work could investigate if such a retrieval strategy (combined with CR), or others such as recalling from another person's perspective, might facilitate face construction further.

Conclusion

This is the first demonstration of an advantage of recalling contextual cues for forensic face construction using modern composite systems in a realistic experimental design (incl. target face unfamiliar at encoding, nominal 24 hr delay to construction, and naming of ensuing composites). Detailed CR, which was used to promote contextual recall of the encoding environment, was successful in increasing memory recall of the face, and thereby facilitating face construction from a recognition system (EvoFIT) and a feature system (PRO-fit)—leading to higher correct naming and likeness ratings of composites for both types of system. Detailed CR was more effective and more consistently effective when constructors had encoded the environment to a greater extent. This mnemonic of the cognitive interview was also effective in combination with an H-CI, with focus on the perceived character of the target face. These two procedures are straightforward, take little time to administer, and if used with eyewitnesses should increase visual identification of offenders. In fact, forensic practitioners in the UK and overseas now regularly use Detailed CR (for recalling scene of crime but not psychological state) with witnesses and victims when constructing a composite with EvoFIT (Frowd et al., 2019). On a final note, it is worth mentioning that, given the similarity of procedures with modern facial composite systems, it seems likely that results would generalise to other feature and recognition systems such as EFIT-V, EFIT-6, FACES 4.0 and Identikit 2000.

References

Alison, L., Doran, B., Long, M.L., Power, N., & Humphrey, A. (2013). The effects of subjective time pressure and individual differences on hypotheses generation and action prioritization in police investigations. *Journal of Experimental Psychology: Applied*, *19*(1), 83-93.

Barak, O., Vakil, E., & Levy, D.A. (2013). Environmental context effects on episodic memory are dependent on retrieval mode and modulated by neuropsychological status. *The Quarterly Journal of Experimental Psychology*, *66*(10), 2008-2022.

Bower, G.H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129-148.

Bower, G.H., & Cohen, P.R. (1982). Emotional influences in memory and thinking: Data and theory. In M.S. Clark & S.T. Fiske (Eds.), *Affect and cognition* (pp. 263-289). Hillsdale, NJ: Erlbaum.

Brown, J.M. (2003). Eyewitness memory for arousing events: Putting things into context. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17*(1)*, 93-106.*

Campeanu, S., Craik, F.I., Backer, K.C., & Alain, C. (2014). Voice reinstatement modulates neural indices of continuous word recognition. *Neuropsychologia*, *62*, 233-244.

Campos, L., & Alonso-Quecuty, M.L. (1999). The cognitive interview: Much more than simply "try again". *Psychology, Crime and Law*, *5*(1-2), 47-59.

Clark, M.S., Milberg, S., & Ross, J. (1983). Arousal cues arousal-related material in memory: Implications for understanding effects of mood on memory. *Journal of Memory and Language*, *22*(6), 633-649.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). *Hil-IsUale, NJ: Lawrence Eribaum Associates.*

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). Improving the reliability of eye-witness identifications: Putting context into context. *Journal of Psychology*, *72*(2), 629-637.

Dando, C.J. (2013). Drawing to remember: External support of older adults' eye-witness performance. *PloS One*, *8*(7), e69937.

Dando, C., Wilcock, R., Milne, R., & Henry, L. (2009). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology*, *23*(5), 698-716.

Davies, G., & Milne, A. (1985). Eyewitness composite construction. A function of mental or physical reinstatement of context. *Criminal Justice and Behaviour*, *12*(2), 209-220.

Davies, G.M. (1983). Forensic face recall: the role of visual and verbal information. In Lloyd-Bostock, S.M.A. and Clifford, B.R. (Eds) *Evaluating Witness Evidence* (pp. 103-123). Wiley, Chichester.

Davies, S. (2009). The effects of stress levels on the construction of facial composites. *Diffusion*, *2*, http://atp.uclan.ac.uk/buddypress/diffusion/?p=1763.

Davis, M.R., McMahon, M., & Greenwood, K.M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*(1), 75-93.

Dodson, C.S., Johnson, M.K., & Schooler, J.W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition*, *25*(2), 129-139.

Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 443-455.

Eich, J.E. (1980). The cue dependent nature of state dependent retrieval. *Memory and Cognition*, *8*(2), 157-173.

Eich, J.E., Weingartner, H., Stillman, R.C., & Gillin, J.C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Memory and Language*, *14*(4), 408-417.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, *8*(1), 101-106.

Emmett, D., Clifford, B.R., & Gwyer, P. (2003). An investigation of the interaction between cognitive style and context reinstatement on memory performance of eyewitnesses. *Personality and Individual Differences*, *34*(8), 1495-1508.

Evans, J.R., Marcon, J.L., & Meissner, C.A. (2009). Cross-racial lineup identification: Assessing the potential benefits of context reinstatement. *Psychology, Crime and Law*, *15*(1), 19-28.

Fisher, R., & Geiselman, R. (1992). *Memory enhancing techniques for investigative interviewing*. Springfield, IL: Charles Thomas Publishers.

Fodarella, C., Frowd, C.D., Warwick, K., Hepton, G., Stone, K., Date, L., & Heard, P. (2017). Adjusting the focus of attention: Helping witnesses to evolve a more identifiable composite. *Forensic Research and Criminology International Journal*, *5*(1), 00143.

Fodarella, C., Kuivaniemi-Smith, H.J., Gawrylowicz, J., & Frowd, C.D. (2015). Detailed procedures for forensic face construction. *Journal of Forensic Practice*, *17*(4), 271-280.

Fodarella, C., Marsh, J.E., Chu, S., Athwal-Kooner, P., & Frowd, C.D. (2019). The importance of context reinstatement for the production of identifiable composite faces from memory: Supplementary experimental material. *Visual Cognition*. (To be uploaded to journal's website on publication.)

Frowd, C.D. (2011). Eyewitnesses and the use and application of cognitive theory. In G. Davey (Ed.) *Introduction to Applied Psychology* (pp. 267-289), BPS Wiley-Blackwell.

Frowd, C.D. (2015). Facial composites and techniques to improve image recognisability. In T. Valentine, & J. Davis (Eds.) *Forensic facial identification: theory and practice of identification from eyewitnesses, composites and cctv* (pp. 43 - 70). Chichester:Wiley-Blackwell. ISBN: 978-1-118-46911-8.

Frowd, C.D. (2017). Facial composite systems: Production of an identifiable face. In: *Face Processing: Systems, Disorders and Cultural Differences*. Nova Science Publishers.

Frowd, C.D., Bruce, V., Ross, D., McIntyre, A., & Hancock, P.J.B. (2007). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition*, *15*(8), 1-31.

Frowd, C.D., Bruce, V., Smith, A., & Hancock, P.J.B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, *14*(3), 276-287.

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L. McLanaghan, S., & Hancock, P.J.B. B. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime and Law*, *11*(1), 33-52.

Frowd, C.D., Erickson, W.B., Lampinen, J.M., Skelton, F.C., McIntyre, A.H., & Hancock, P.J.B. (2015). A decade of evolving composite techniques: Regression- and meta-analysis. *Journal of Forensic Practice*, *17*(4), 319-334.

Frowd, C.D., & Fields, S. (2011). Verbalisation effects in facial composite production. *Psychology, Crime and Law*, *17*(8), 731-744.

Frowd, C.D., Hancock, P.J.B., Bruce, V., McIntyre, A.H., Pitchford, M., Atkins, R., Webster, A., Pollard, J., Hunt, B., Price, E., & Morgan, S. (2010, September). Giving crime the'evo': catching criminals using EvoFIT facial composites. In *2010 International Conference on Emerging Security Technologies* (pp. 36-43). IEEE.

Frowd, C.D., Jones, S., Fodarella, C., Skelton, F.C., Fields, S., Williams, A., Marsh, J., Thorley, R., Nelson, L., Greenwood, L., Date, L., Kearley, K., McIntyre, A., & Hancock, P.J.B. (2014). Configural and featural information in facial-composite images. *Science and Justice, 54*(3), 215-227.

Frowd, C.D., Nelson, L., Skelton F.C., Noyce, R., Atkins, R., Heard, P., Morgan, D., Fields, S., Henry, J., McIntyre, A., & Hancock, P.J.B. (2012). Interviewing techniques for Darwinian facial composite systems. *Applied Cognitive Psychology*, *26*(4), 576-584.

Frowd, C.D., Pitchford, M., Skelton, F.C., Petkovic, A., Prosser, C., & Coates, B. (2012). Catching even more offenders with EvoFIT facial composites. In A. Stoica, D. Zarzhitsky, G. Howells, C. Frowd, K. McDonald-Maier, A. Erdogan, and T. Arslan (Eds.) *IEEE Proceedings of 2012 Third International Conference on Emerging Security Technologies* (pp. 20 - 26). DOI 10.1109/EST.2012.26.

Frowd, C.D., Portch, E., Killeen, A., Mullen, L., Martin, A.J., & Hancock, P.J.B. (2019). EvoFIT facial composite images: a detailed assessment of impact on forensic practitioners, police investigators, victims, witnesses, offenders and the media. In A. Stoica, G. Howells, K. McDonald-Maier, A. Erdogan, and T. Arslan (Eds.) *IEEE Proceedings of 2019 Eighth International Conference on Emerging Security Technologies*, July, University of Essex.

Frowd, C.D., Skelton, F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., McIntyre, A.H., & Hanock, P.J.B. (2012). Recovering Faces From Memory: The Distracting Influence of External Facial Features. *Journal of Experimental Psychology: Applied, 18*(2), 224-238.

Gawrylowicz, J., Gabbert, F., Carson, D., Lindsay, W.R., & Hancock, P.J.B (2012). Holistic versus featural facial composite systems for people with mild intellectual disabilities. *Applied Cognitive Psychology*, *26*(5), 716-720.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P., & Holland, H.L. (1985). Eyewitness memory enhancement in the police interview: cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*(2), 401-412.

Glenberg, A.M. (1997). What memory is for. *Behavioral & Brain Sciences, 20*(1), 1-55.

Godden, D.R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: Land and underwater. *British Journal of Psychology*, *66*(3), 325-331.

Goldstein, A.G., & Mackenberg, E.J. (1966). Recognition of human faces from isolated facial features: A developmental study. *Psychonomic Science*, *6*(4), 149-150.

Hasel, L.E., & Wells, G.L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and human Behavior*, *31*(2), 193-207.

Hammond, L., Wagstaff, G.F., & Cole, J. (2006). Facilitating eyewitness memory in adults and children with context reinstatement and focused meditation. *Journal of Investigative Psychology and Offender Profiling*, *3*(2), 117-130.

Hanczakowski, M., Zawadzka, K., & Coote, L. (2014). Context reinstatement in recognition: Memory and beyond. *Journal of Memory and Language*, *72*, 85-97.

Hockley, W.E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1412-1429.

Kehn, A., Renken, M.D., Gray, J. M., & Nunez, N.L. (2014). Developmental trends in the process of constructing own-and other-race facial composites. *The Journal of psychology*, *148*(3), 287-304.

Koen, J.D., Aly, M., Wang, W.C., & Yonelinas, A.P. (2013). Examining the causes of memory strength variability: Recollection, attention failure, or encoding variability? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1726-1741.

Laughery, K.R., Duval, C., & Wogalter, M.S. (1986). Dynamics of facial recall. In A.W. Young (Eds.) *Aspects of face processing* (pp. 373-387). Dordrecht: Springer Netherlands.

Levy, D.A., Rabinyan, E., & Vakil, E. (2008). Forgotten but not gone: Context effects on recognition do not require explicit memory for context. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1620-1628.

Macken, W.J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 153-161.

Malpass, R.S., & Devine, P.G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*(4), 482-489.

Martin, A.J., Hancock, P.J.B., & Frowd, C.D. (2017, September). Breathe, relax and remember: An investigation into how focused breathing can improve identification of EvoFIT facial composites. In *2017 Seventh International Conference on Emerging Security Technologies (EST)* (pp.79-84). IEEE.

Martin, A.J., Hancock, P.J.B., Frowd, C.D., Heard, P., Gaskin, E., Ford, C., & Hewitt, T. (2018). EvoFIT composite face construction via practitioner interviewing and a witness-administered protocol. In G. Howells et al. (Eds.) *Proceedings of 12th NASA / ESA Conference on Adaptive Hardware and Systems*, 6th - 9th August, University of Edinburgh, Edinburgh, UK.

Memon, A., & Bruce, V. (1995). Context effects in episodic studies of verbal and facial memory: A review. *Current Psychological Research and Reviews*, *4*(4), 349-369.

Memon, A., Meissner, C.A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*(4), 340-372.

Milne, R., & Bull, R. (2002). Back to basics: A componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology*, *16*(7), 743-753.

Murnane, K., & Phelps, M.P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 882-894.

Ness, H., & Bruce, V. (2006). An investigation into the use of CCTV footage to improve likeness in facial composites. *European Association of Psychology and Law Conference*, 27-30 June 2006, University of Liverpool, Liverpool, UK.

Neyeloff, J.L., Fuchs, S.C., & Moreira, L.B. (2012). Meta-analyses and forest plots using a microsoft excel spreadsheet: Step-by-step guide focusing on descriptive data analysis. *BMC research notes*, *5*(1), 52.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, *30*(6), 755-764.

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are pre-schoolers sensitive to configural information in faces? *Developmental Science*, *9*(3), 270-277.

Rainis, N. (2001). Semantic contexts and face recognition. *Applied Cognitive Psychology*, *15*(2), 173-186.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*(2), 139-156.

Smith, S.M. (2013). Effects of environmental context on human memory. In *The SAGE handbook of applied memory*, *2*, 162-182.

Smith, S.M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(5), 460-471.

Smith, S.M., & Manzano, I. (2010). Video context dependent recall. *Behavior Research Methods*, *42*(1), 292-301.

Smith, S.M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin and Review*, *8*(2), 203-220.

Thomson, D.M., Robertson, S.L., & Vogt, R. (1982). Person recognition: The effect of context. *Human Learning*, *1*(1), 37-54.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory, *Psychological Review*, *80*(5), 359-380.

Wagstaff, G.F. (1982). Context effects in eyewitness reports. Paper presented at the *Law and Psychology Conference*. Swansea, Wales, July.

Wagstaff, G.F., Cole, J., Wheatcroft, J., Marshall, M., & Barsby, I. (2007). A componential approach to hypnotic memory facilitation: Focused meditation, context reinstatement and eye movements. *Contemporary Hypnosis*, *24*(3), 97-108.

Wagstaff, G.F., Wheatcroft, J.M., Caddick, A.M., Kirby, L.J., & Lamont, E. (2011). Enhancing witness memory with techniques derived from hypnotic investigative interviewing: Focused meditation, eye-closure, and context reinstatement. International *Journal of Clinical and Experimental Hypnosis*, *59*(2), 146-164.

Wells, G.L., Memon, A., & Penrod, S.D. (2007). Eyewitness evidence: improving its probative value. *Psychological Sciences in the Public Interest*, *7*(2), 45-75.

Wong, C.K., & Read, J.D. (2011). Positive and negative effects of physical context reinstatement on eyewitness recall and identification. *Applied Cognitive Psychology*, *25*(1), 2-11.

3

GENERAL DISCUSSION

The aim of this thesis was to improve the effectiveness of facial composites by using the context that was present during time of encoding as a retrieval cue to facilitate eyewitness memory of the to-be-remembered target face. Facial composite systems have improved dramatically over the past years (for a review, see Frowd, 2017) and there are various manipulations available to increase the effectiveness of composites, implemented either before composite construction or to completed images (see Chapter 1: Literature Review for a more in-depth review). However, as composites are created from an eyewitness's memory, a certain degree of error is naturally transmitted, and it is therefore pertinent to create further techniques to facilitate face construction. The current research aimed to do this by using contextual cues (see Chapter 1: Introduction to context reinstatement effects). The empirical literature has consistently shown that by reinstating the context—both physically and mentally—in which memory encoding occurred, contextual information can act as a retrieval cue leading to increased memory recall or recognition. Context reinstatement (CR) has been used effectively to increase verbal (e.g., Brinegar et al., 2013; Godden & Baddeley, 1975; Smith, 1979), facial (e.g., Evans et al., 2009; Thomson et al., 1982; Wagstaff, 1982) and eyewitness memory (e.g., Dietze et al., 2013; Memon & Bruce, 1995; Wagstaff et al., 2007; Wong & Read, 2011), and is used in the real world as part of the CI to interview eyewitnesses and victims of crime (Geiselman et al., 1985; Wells et al., 2007). Yet, it seems to have been scarcely utilised to also improve the effectiveness of facial composites. The current research, encompassed as a formal research paper in Chapter 2, aimed to explore whether the CR technique can be effective with contemporary systems to inform police practice. This chapter will discuss the main findings, as well as offer suggestions for future research.

Using context cues to facilitate composite identifiability

Davies and Milne (1985) published what appears to be the first paper to apply CR techniques prior to constructing a composite, finding that a Physical CR improved composite effectiveness, but that a Mental CR led to even-better quality (more recognisable) Photofit composites. In a conference paper (Ness & Bruce, 2006) and an unpublished experiment (Ness et al., 2004), a physical CR benefit on ensuing PRO-fit composites was also found using video-recorded material of the original environment. The current research aimed to replicate and extend past findings of a CR benefit using contemporary systems for facial composite construction. As described in Chapter 2, results from Experiment 1 revealed that physically reinstating the context led to improved likeness ratings of both EvoFIT and PRO-fit composites, compared to the control condition Minimal CR, consisting of only mentally visualising and recalling the face. As the Physical CR condition consisted of the same interviewing procedure, it becomes clear that the sole act of returning to the original encoding environment helped in creating composites that are closer matches to the target face, a finding which supports Ness and Bruce's (2006) research.

Even higher likeness ratings, however, were given to composites constructed via Detailed CR[6], with which participant-witnesses mentally visualised and recalled the environmental as well as internal context in detail prior to recalling the face. With regard to identification, Detailed CR also significantly increased correct naming (cf. Minimal and Physical CR), which indicates that participants' memory was facilitated in this condition. As participants were first involved in recalling the environmental context as well as their mood and feelings, it appears as though this information acted as associative retrieval cues, facilitating access to facial memory for the target (Tulving & Thompson, 1973). Physical CR did not have this effect to the same extent: although likeness ratings increased compared to Minimal, this improvement was not strong enough to also facilitate naming rates. This outcome may be explained by the fact that the interviewing procedure prior to composite construction did not differ between Minimal and Physical conditions—that is, participants only actively recalled the target face, but did not engage in an intensive memory recall as they did in Detailed CR. Therefore, the environmental context as

[6] Please note that Detailed CR is equivalent to Mental CR from Davies and Milne's (1985) experiment. As the police prior to composite construction already employ a minimal Mental CR interview, consisting of 'thinking back' as well as visualising and recalling the face, the current thesis chose to distinguish between Minimal and Detailed CR.

a physical retrieval cue in itself was not strong enough to aid participants' memory to the point of improving ensuing composites sufficiently to be more recognisable: they were only better-rated when compared to the original target, an improvement which is arguably of little applied benefit. It may be that with Detailed CR, participants were guided to engage in an active recall of the environmental and internal context, whereas in Physical CR, it was down to the participants themselves whether or not to engage with the physically reinstated context that was present at the time of construction.

A way of controlling this issue would be to specifically ask participants to attend to the environment, either by asking them to verbalise what they see, or to simply look around prior to constructing the face. Also, the same interviewing procedure as utilised in Detailed CR could be employed during Physical CR, that is, to recall the original environmental and internal context that was present during the time of encoding. This procedure would be equivalent to that in Davies and Milne's (1985) experiment that in fact appeared to be more effective than employing a Detailed CR in a different room—however, only by 3%. Although this technique could have been explored further in subsequent experiments, it was decided not to investigate it in more detail for two reasons. A Physical CR would pose practical problems for the police by having to use valuable time and cost to return to the crime scene, two factors that are usually scarce within policing. A more pressing issue would perhaps be ethical considerations for the witness and especially the victim. Having to return to the crime scene could be considerably more distressing and traumatic than having to mentally visualise and recall the environment. This distress could in turn reduce the accuracy of an eyewitness's memory and thereby negatively impact on their ability to create a good-quality composite (Davies, 2009; Hancock et al., 2011). For these reasons, as well as the fact that a Detailed CR is a straightforward and convenient procedure to implement, with fewer ethical issues, the last five experiments of this thesis only employed a Detailed CR.

Surprisingly, Detailed CR was not effective in the Supplementary experiment; in fact, likeness ratings of EvoFIT composites decreased significantly, with no reliable change in naming rates. PRO-fit naming and rating means were in the expected direction; that is, they increased, but this difference was found to be non-significant.

Given the apparent discrepancy in results between the first experiment and the Supplementary experiment for the context manipulation, it was posited that differences might be due to two possible (not necessarily mutually-exclusive) reasons. First, it was thought that inconsistencies may have derived from the EvoFIT system, since results for PRO-fit composites were in the expected direction; the fact that this was non-significant may be due to the very low PRO-fit naming rates having caused a floor effect. The EvoFIT system that was used during the Supplementary experiment involved an older, less-effective selection procedure that may have led to more variable results: faces shown in face arrays were selected based on their overall likeness to the original target face. However, a newer technique has been developed which not only leads to an improved likeness but also to more consistent findings: to select for the *upper facial half* (Fodarella et al., 2017). This technique may be effective by ensuring that constructors do not neglect the upper facial region during selection, a region that is important for identification of familiar faces (e.g., Goldstein & Mackenberg, 1966; O'Donnell & Bruce, 2001; Pellicano et al., 2006) as well as for facial composites (Ellis et al., 1980; Frowd et al., 2007, 2011; Laughery et al., 1986). As such, if constructors focus on this region during composite construction, they are more likely to create an upper facial likeness that is accurate, thereby leading to a more-identifiable ensuing composite. The inconsistencies derived through the old selection procedure are likely to stem from constructors not sufficiently focusing on the upper region. Therefore, using this new, improved selection procedure in subsequent experiments of the current thesis (Experiments 3-5) may also ensure that CR effects are more-consistently detected.

Second, inconsistencies may be due to participant effects; for example, constructors could have been particularly fatigued or anxious in the Supplementary experiment during face construction, perhaps due to demands of university study (esp. as many constructions took place during an exam period); as such, it is conceivable that participants in the Supplementary experiment might have simply paid little attention to the environment. This would be in line with the notion that emotion or anxiety can cause a narrowing of attention, leading to reduced cue utilisation (Easterbrook, 1959). In the current research, poor encoding of the environment would result in inadequate recall of contextual information which in turn would not facilitate memory for composite construction. In addition, while the Supplementary experiment did not find a benefit of Detailed CR (cf. Minimal CR) by naming, the manipulation did not make it reliably worse, with differences presumably attributable to noise; this indicates a potentially weak and/or variable effect. This proposal would square with the notion that the environment may simply not have been encoded sufficiently.

Experiment 2 investigated the possibility of poor context encoding by repeating the experiment at another time, avoiding exam periods, as well as using an improved EvoFIT selection procedure in an attempt to ensure the consistency of findings and to mirror current forensic practice. Thereby, Experiment 2 employed an identical design as the Supplementary experiment and found that Detailed CR was successful in increasing correct naming of composites, overall. The interaction effect revealed, however, that this was the case for EvoFIT but not for PRO-fit composites. PRO-fit ratings suggested an improved likenesses following Detailed CR (cf. Minimal), but presumably this improved likeness was not sufficient to also aid correct naming; the fact that PRO-fit composites were poorly named may have influenced this outcome (by reduced experimental power).

Since findings were still inconsistent following Experiment 2, the notion of poor or insufficient attention to the environment was investigated more-thoroughly in the subsequent experiments. The environment may have been encoded poorly—in line with the theory of environmental suppression (Glenberg, 1997)—as some participants were preoccupied with the conceptual processing of encoding the target face and as a result did not pay attention to the environment. This explanation seems plausible as the researcher met participants outside of the testing environment (campus café), walked in whilst engaging in informal conversation and chose where the participant was to be seated to begin face encoding. Participants may have simply followed the researcher without paying attention to the environment, and were subsequently preoccupied with the testing task.

Therefore, Experiment 3 manipulated attention to the environment: to be either intentional, by asking participants to sufficiently attend to the environment, or incidental, by not providing such instructions. Indeed, findings were in favour of the proposed theory that inconsistencies in previous experiments derived from insufficient encoding of the environment. First of all, there was no significant difference for Minimal CR conditions across attention type. This result is sensible as the environment was not actively utilised prior to composite construction, and so it was not expected to differ whether or not constructors' attention was directed to the environment. This finding also implies that the sole encoding of the environment is not sufficient to later provide access to the target memory, but only the act of mentally visualising and recalling it.

Moreover, Detailed CR only reliably increased composite naming following intentional encoding. Following incidental encoding, Detailed CR was not effective, thereby mirroring findings from the Supplementary experiment. Rating data further revealed that composites constructed via Detailed CR in incidental conditions, albeit not better recognised, were actually better-rated compared to those from Minimal CR. This implies that although participant-witnesses were not asked to encode the environment, they must have done this regardless to some extent, thereby in part benefitting from context cues to facilitate composite construction. This improved resemblance was only effective in increasing likeness ratings, not correct naming, possibly due to the fact that the environment was not encoded to the same extent as in intentional conditions. This shows that since Detailed CR was also successful in Experiments 1 and 2, participants may have paid sufficient attention to the environment without having been specifically encouraged to do so, whilst in the Supplementary experiment and in the incidental condition in Experiment 3, this was not the case. These results through intentional encoding are also somewhat in line with past research that has demonstrated context cues to be more effective in facilitating memory when an interactive encoding between the environmental context and the to-be-remembered-target was promoted (see Hanczakowski et al., 2015; Hockley, 2008). The current author acknowledges that Experiment 3 did not implement an actual *interactive* encoding between the participant and the environment, and it would be fruitful to explore this in future research, thereby also mirroring a more realistic scenario than used in the current work. As the environment in the current research is perhaps more 'disconnected' from the target face than it would be in real life, and participants followed the researcher into the testing environment, one would theorise that the issue regarding lack of attention to the environment would not occur for real eyewitnesses.

To guide participants' attention to the environment more-naturally, in Experiment 4 participants were invited to enter the testing environment on their own and navigate to a designated chair. This procedure aimed to promote sufficient attentiveness to the environment incidentally rather than intentionally, arguably mirroring a more realistic way in which witnesses may encode an environmental context. Using this method, Detailed CR was successful in significantly increasing correct naming over Minimal CR, revealing that this novel procedure did indeed promote a greater environmental attention, without additional prompts. Despite the fact that the effect size for Detailed CR following intentional encoding in the previous experiment was—unsurprisingly perhaps—larger (d = 1.40), the advantage for Detailed CR following incidental encoding in Experiment 4 was nonetheless large (d = 1.08).

Therefore, it would appear as though the stability of the effect is mediated by the extent to which the environment is encoded. As incidental conditions led to the most variable results, the current work further aimed to quantify the magnitude and reliability of the context effect in those conditions. Thus, meta-analyses were conducted of Experiments 1 to 3, as well as the Supplementary experiment, and using the more forensically relevant composite naming as DV. An overall reliable benefit of Detailed CR (cf. Minimal CR) was found for both composite systems, with a small effect size, and for PRO-fit and EvoFIT systems separately, also with small effect sizes. The analyses further revealed that the positive effect is reliable and generalisable to both other participants and other targets (items). Experiments in the thesis also intentionally used a variety of target images to increase confidence in generalisability of the effect: both male and female targets, and from different TV soaps (Coronation Street, EastEnders) and type of sport (Football, Rugby).

As eyewitnesses or victims of real crimes may not encode the environment intentionally, it is promising that participant-witnesses benefitted from context cues to produce better-quality composites even under incidental conditions. This was evidenced by the meta-analyses as well as by findings from Experiment 4, wherein attention to the environment was increased naturally, but remained incidental. This should therefore reduce concerns regarding stability of the context effect as, in real life, witnesses are more likely to encode the environment sufficiently, similar to how constructors achieved this in Experiment 4. It is also important to highlight that even when Detailed CR was not successful, it did not have a detrimental effect on correct naming of composites—that is, there was no reliable reduction in correct naming. Incorrect naming was also not affected as indicated by findings from the meta-analysis, suggesting that the manipulation should not lead to a change in misidentification of composites. Therefore, it can be argued that there is no disadvantage in utilising this interviewing mnemonic within the context of a real world face construction.

In order to appropriately inform law enforcement, the current work also explored whether Detailed CR would be effective using a novel method of composite construction whereby constructors produce a facial likeness on their own (e.g., Martin et al., 2018). This new version of EvoFIT software was created (see www.EvoFIT.co.uk) for less serious crime, such as minor theft or anti-social behaviour, where police officers do not have the time to work with a witness to produce a composite using the traditional face-to-face method (e.g., Alison et al., 2013). Using the novel approach, witnesses are shown on-screen instructions to guide them through the composite construction procedure, as far as possible to ensure that this procedure is equivalent to how a composite would be created with a facial imaging specialist. Experiment 4 revealed that although overall, self-administered composites were not as identifiable as those constructed using the traditional face-to-face procedure, Detailed CR was effective for both methods. This implies that constructors may still benefit from traditional interaction wherein verbal guidance may better-aid their understanding of the process than written instructions, potentially as the former could be argued to be more engaging, less taxing and to require less concentration (see Kalyuga, 2012). Also, the face-to-face construction method enables constructors to ask questions in order to clarify any aspect of the procedure, which may in turn have a positive effect on ensuing composites. It is promising though that the Detailed CR technique was also effective using the new self-administered method, suggesting that constructors did genuinely engage in a self-administered recall of environment context to facilitate their subsequent construction. As self-administered composites were of lower quality, it is a positive outcome that Detailed CR would be a quick and simple instruction to include in order to increase performance. In Experiment 4, Detailed CR increased correct naming nearly to that of the face-to-face method without Detailed CR (M = 50.0 and 54.7 respectively).

The procedure of Detailed CR employed in Experiments 1 to 3 as well as the Supplementary experiment involved free recall of not only the environmental context, but also the psychological context, that is, of the constructor's moods and feelings at the time of encoding. In Experiment 4, Detailed CR was reduced to only include recall of the environment, and results suggest that this in itself reliably facilitated composite construction. This is a valuable finding for police practice as eyewitnesses, and especially victims of serious crime, may find it traumatic having to recall their emotional state at the time of the crime. It would be expected that their anxiety levels would increase as a result, which in turn would inhibit facial composite construction (Davies, 2009). Similarly, more identifiable composites are also produced when procedures are used to intentionally reduce anxiety levels (Martin et al., 2017). It is therefore promising that reference to the witness's psychological state can be omitted from the interviewing procedure whilst maintaining the facilitating effect of context on face production. Experiment 4 showed that this benefit (i.e., without psychological recall) was possible using both the traditional face-to-face procedure and the novel 'self-administered' method.

It is unclear as to why recall of the internal context is not necessary to facilitate access to the memory of the face. In the current work, participants generally spent more time recalling the environmental rather than their internal context, and so this may be indicative that the environment is more helpful in aiding memory access. It may also be that although in Experiment 4 participants were not asked to verbalise their internal context at the time of face encoding, the act of recalling the environment may have automatically triggered memory of their moods and feelings at the time, as would be implied with the encoding specificity principle (Tulving & Thompson, 1973). It would have been theoretically interesting to have asked participants whether this was the case. Also, the author does acknowledge that a more empirical comparison of context types could have been conducted, and it would be sensible to consider this in future work. More-specifically, effects on composite construction from recalling (i) the environmental context alone, (ii) the psychological context alone, (iii) both the environmental and psychological context. Whilst it seems theoretically and practically valuable to explore this in more detail, it appears unlikely that additional or sole recall of psychological state would be effective when interviewing witnesses and victims in real life.

Further analyses were conducted to explore the underlying mechanism that leads to the effectiveness of Detailed CR in facilitating composite construction. The current research pooled experiments together in the meta-analysis to increase power and thus the likelihood of detecting a significant difference. Meta-analyses indicated that Detailed CR lead to a reliable increase in constructors' face recall. More-specifically, this included, recall of the target's individual facial features as well as a more holistic memory recall of the target face, such as personality attributes and characteristics. Meta-analyses only emerged reliable for those experiments where Detailed CR was successful, namely, Experiments 1, 2 and 4. The significant effect ceased to exist once data from Experiments 3 and the Supplementary experiment were included, both of which had not led constructors to produce more accurate composites following Detailed CR. Whilst the current findings contrast those of Davies and Milne's (1985), who found no increase in total face recall as a result of use of context cues, this discrepancy could be explained by a difference in how recall was elicited: the current research employed a free recall procedure (for both context and the face), whereas Davies and Milne (1985) employed a standard interview, using a question-and-answer type procedure, and so recall was not entirely free. Further, fewer descriptors were elicited than in the current experiments, whereby the measure may have been less sensitive in detecting any differences in face recall across conditions. It should be noted here, that the current work did not distinguish between correct and incorrect face recall, and so it may be worthwhile to investigate in future whether recall is not only less complete, but also less accurate without the use of context cues. Such follow-up work might assess constructors' descriptions for accuracy (or inaccuracy), presumably with the finding that a more accurate recall would lead to more accurate composites.

Analyses of the current face recall data alone cannot indicate whether Detailed CR merely facilitates face recall or also additionally face recognition, two indexes of facial memory performance that are not necessarily mutually-exclusive. Therefore, Experiment 5 reversed the order of recall for environmental context and face to explore whether a facilitating effect would remain. If recognition was being improved through the use of context cues, this reversed recall should continue to be effective to some extent during composite construction by increasing recognition of full faces and facial features shown to the participant-constructors. Contrary to past findings, wherein cues associated with the target face enabled better face recognition (Shapiro & Penrod, 1986), the current experiment did not find this to be the case. There was also no apparent trend in the data, one which may have indicated an issue with experimental power; here p-values for CR and the interaction were high ($p > .5$)—this was also in spite of collecting additional (cued naming) data. Therefore, it would seem that Detailed CR only works by an increase in facial recall rather than recognition.

A further issue in using this reversed recall order may have been the fact that constructors first engaged in face recall that is not as complete (and potentially not as accurate) as it would have been following the contextual recall (as evidenced by findings from the meta-analysis). This situation may be similar to findings that indicates witnesses suffer from an increased false confidence in their testimony or recall once verbalised (Odinot, Wolters, & van Giezen, 2012; Shaw & McClure, 1996). Thus, in the current work, there may have been an issue with constructors abiding by their original, relatively incomplete—and perhaps more inaccurate—face recall. Although the EvoFIT system arguably aims to promote face recognition during the majority of the construction process, more-detailed manipulation of individual facial features, for example, the eyes which are important for subsequent identification, still occurs through face recall. During this process, it is possible that the witness's initial face recall may have interfered with the construction of a better-quality composite.

Combining CR with further interviewing mnemonics

Since the Detailed CR interview was effective in increasing composite quality in the experiments of the current thesis, it was also aimed to improve composites further by combining Detailed CR with other interviewing mnemonics known to facilitate memory of facial composites. One such mnemonic is the 'Holistic' Cognitive Interview (H-CI). Prior to composite construction, witnesses are commonly interviewed via a Cognitive Interview (CI) which invites them to mentally visualise the target face and to freely recall as many details about it as possible (see Frowd et al., 2011). In contrast, the H-CI has been found to increase composite effectiveness over the standard CI for face recall. With the H-CI, witnesses first engage in the standard CI, but are subsequently asked to reflect silently upon on the personality of the face for 60 seconds, before rating personality traits (e.g., intelligence, friendliness, kindness) on a three-point Likert scale, thus making whole-face judgments (Frowd et al., 2008, 2013, 2015; Frowd, Nelson, et al., 2012).

Experiment 2 as well as the Supplementary experiment combined the H-CI with Detailed CR. In both experiments, the H-CI overall significantly increased naming and likeness ratings for composites from both systems (cf. CI). The rating data also revealed significantly greater improvement from CI to H-CI for composites from the feature system PRO-fit than the recognition system EvoFIT. This may be due to feature systems simply having a greater scope for improvement since their quality is inferior to recognition systems (Frowd et al., 2015). The findings are in line with past research (e.g., Frowd et al., 2008, 2013, 2015; Frowd, Nelson, et al., 2012), and this improvement through H-CI is possibly due to a greater, more-in-depth involvement and analysis of the target face (in order to assign personality attributes) than with the CI. Also, H-CI is more in line with how we perceive and remember faces, that is, as a whole (e.g., Richler et al., 2011; Young et al., 1987). Whilst the standard CI promotes face description through feature recall—likely to interfere with recognition (Dodson et al., 1997)—the H-CI is theorised to promote holistic facial recall (e.g., Davies & Oldman, 1999)—likely to improve recognition (Davies & Oldman, 1999)—thereby increasing recall and in turn also facial composite likeness. Further to this, internal and external features as well as age and distinctiveness improve following the H-CI (Frowd et al., 2008), factors that not only facilitate face recognition (Ellis et al., 1979), but also composite recognition (Frowd et al., 2008).

In Experiment 2, it was also apparent that the H-CI procedure was effective in combination with the Detailed CR technique, leading to the overall highest naming and likeness ratings compared to other conditions. These two techniques seem to guide constructors' attention to different aspects of the face and this is perhaps why they are successful in combination. It is well-established that the H-CI increases attention to the face as whole as well as to its central parts (e.g., Frowd et al., 2008) rather than featural processing of the face (e.g., Davies & Oldman, 1999). This is because the focus is on the personality of the face—an aspect which requires memory of and focus on the entire face as opposed to its individual features (Mayes, Meudell, & Neary, 1980; Patterson & Baddeley, 1977). In contrast, Detailed CR appears to aid memory access to individual facial features, as evidenced by the meta-analysis of constructors' face recall. If facial memory is facilitated by providing better access to memory of the whole face as well as its individual features, it is not surprising that these techniques would work well in combination, resulting in a more-accurate face. During composite construction, guiding a witness's attention to the central part of the face assists in creating a composite of better effectiveness (e.g., Frowd et al., 2008). If featural memory is also facilitated, it is likely to assist a witness in better discriminating between facial features, such as eyes and brows. This upper facial region is an important factor in familiar face recognition (O'Donnell & Bruce, 2001)—in the current work, pertinent to facial composite naming—and improved memory for this region would likely enable the witness to construct a better likeness (Fodarella et al., 2017), and thereby increasing ensuing identification (Ellis et al., 1980).

In Experiment 3, Detailed CR was combined with a different interviewing mnemonic, extensive recall, with the aim of aiding facial memory further. This technique is based on the notion that memory recall is increased by engaging in multiple retrieval efforts instead of stopping after only the first attempt at memory search (Fisher, & Geiselman, 1992). The current work implemented this interviewing mnemonic in Experiment 3: participant-witnesses first provided a free recall of the environmental and psychological context, and were then asked further open-ended, probing questions. It was theorised that this would result in greater memory recall of the contextual information and thereby facilitate subsequent face recall and composite construction. The data, however, do not suggest that multiple recall attempts are effective; there was no significant increase in either naming nor ratings compared to Detailed CR alone. It is not entirely clear as to why this additional technique was not successful. Participant-witnesses were able to recall additional details when further probed, showing that the extensive technique elicited more information compared to their initial, free recall. One potential reason for the null effect is that the additional information was merely not effective in aiding facial memory any further, perhaps because it was not as relevant as the initial free recall. A further reason may be that this additional information was already mentally visualised during the free recall stage, but it was not verbalised to the experimenter, despite the participant being asked to recall as many details as possible. The composite experimenter did not specifically ask witnesses whether they felt they mentally reconstructed the environment to a greater extent following the extensive recall or simply verbalised more, and it would be beneficial to clarify this in future experiments.

Another potential reason for the ineffectiveness of the extensive recall technique may be linked to the retrieval-induced forgetting literature. This phenomenon indicates that the repeated retrieval of a subset of information reduces recall of non-practised information (e.g., MacLeod, 2002; Shaw, Bjork, & Handal, 1995). Whilst in the current work, the repeated retrieval of context cues did not reliably reduce composites quality (cf. Detailed CR and Minimal CR), it may have resulted in the offset of any potential gains.

Finally, findings from the current work also seem to be somewhat related to Campos and Alonso-Quecuty's (1999) results, wherein other retrieval techniques—namely, CR, recall in a different order, change perspective, and other strategies—were more successful in facilitating correct recall than multiple retrieval efforts (four cycles of "try again"). Whilst instructions given to participant-witnesses in Experiment 3 were not to "try again" (as in Campos & Alonso-Quecuty, 1999), the cued-based recall questions here may be less effective than other retrieval techniques. It would be reasonable to consider combining other retrieval mnemonics with Detailed CR in future work, such as asking participants to recall in a different spatial order, or from another person's perspective (see Fisher & Geiselman, 1992). Similarly, past research has combined MCR with a focused meditation (FM; where individuals listen to a 1.5 minute audiotape and engage in focused breathing) and revealed an additive effect in increasing correct eyewitness recall (Wagstaff et al., 2007, 2011). A focused breathing technique (where constructors paused for about 10 seconds, closing their eyes and concentrating on their breathing) has already found to be effective in increasing recognisability of ensuing composites (Martin et al., 2017), and so combining this breathing technique or FM with Detailed CR may be even more effective in improving ensuing composite likeness. Perhaps the interviewing techniques discussed here (recall in a different spatial order, change perspective, focused breathing, FM) may be more powerful (cf. Extensive recall used in Experiment 3) in combination with Detailed CR to facilitate facial composite construction.

Differences between composite systems

The first two experiments as well as the Supplementary experiment from the current research utilised both a typical holistic (EvoFIT) and a typical feature system (PRO-fit). These two types of systems were included primarily to explore whether manipulations would be effective for both systems. Although a comparison of the systems was not the main aim, they have been statistically compared in the thesis really to ensure replication of past findings. Compared to PRO-fit, EvoFIT composites were significantly better-named in Experiments 1, 2 and the Supplementary experiment, as well as higher-rated in Experiment 1 and the Supplementary experiment. This was expected: across the literature, PRO-fit composites are less effective than EvoFIT composites (e.g., see Frowd et al., 2015 for a recent meta-analysis). This is possibly due to the fact that feature systems—such as PRO-fit—require a more-detailed face recall than recognition systems—such as EvoFIT—which facilitate face construction through face recognition rather than recall. As face recognition is a less challenging task than facial recall (Davis et al., 1961), the EvoFIT system enables constructors to create composites that are of better-quality than the PRO-fit system. Also, it could be argued that the completed EvoFIT face images appear more realistic than PRO-fit composites, and this may have an effect on identification rates.

The underlying mechanism that drives Detailed CR to effectively improve composite quality seems to be an increase in face recall. The increased face recall—in those experiments where Detailed CR was successful in facilitating face construction—was effective in improving composite quality for both types of systems. It is perhaps more plausible that PRO-fit, which requires a more-detailed recall of the face as part of the construction procedure, would benefit especially from an increase in constructors' recall. However, it may at first seem counterintuitive that EvoFIT, being a recognition system, would also benefit from an increase in face recall. Nonetheless, an improved face recall is likely to also lead to an improved recognition of the face. Further, parts of the EvoFIT procedure require face recall, namely, when manipulating individual facial features towards the end of the procedure, and so constructors' should benefit from improved recall of the facial features. Finally, as face recall also included personality attributes and characteristics, it is not surprising that this holistic information aided EvoFIT composite construction.

The final experiments no longer utilised PRO-fit due to its very low naming rates in the initial experiments as well as the fact that police in the UK no longer use feature systems—unlike at the beginning of Experiments 1 and 2, wherein feature systems were still in use and research would have informed police practice. In the US, feature systems are still utilised, such as Identikit 2000 and FACES 4.0, and so it would be sensible not only to replicate current findings but also to further investigate context effects using composite systems as used in the US.

Future experiments

Whilst the current research provides evidence that a Detailed CR can be used to promote composite identification rates for both holistic and feature systems, more research could be conducted to explore how the effect might be mediated by other factors. Exploring factors that influence CR effects is not only interesting with regard to psychological theory, but also crucial to inform forensic practitioners as to when and how to use the Detailed CR technique. Whilst this chapter has already highlighted potential future work in relevant sections above, other areas which appear to be worthy of further explorations are outlined below.

Aside from encoding of the environment, a further potential factor that may have contributed to the inconsistencies in findings is the retention interval deployed within studies. Context cues are generally more effective when the retention interval is longer (e.g,. 2 weeks rather than 2 days) and memory thereby worse (Cutler et al., 1987). Davies and Milne's (1985) experiment which found that context cues were effective in improving the resulting composite, constructed one after a one-week-delay, whilst the current experiments utilised a retention interval of 24 hours. Thus, it may be that participant memory was not poor enough in *all* cases, but only in *some* cases, in order to benefit from context cues to provide access to memory of the target face, thereby contributing to the inconsistency of findings. This hypothesis could be further explored in future research by manipulating the retention interval to explore how context effects are mediated, and/or by reducing encoding time of the target face. It would be expected that the context effect would become stronger with an increasing retention interval or shorter encoding time.

Another factor that may influence the effectiveness of the Detailed CR technique could be related to the familiarity of the environment. In all experiments, the environment was chosen to be unfamiliar to face constructors, and so, subsequent experiments could investigate whether CR effectiveness would be compromised when the environment is familiar. This would inform practice for crimes, such as distraction burglaries, which do occur in a victim's home, or other crimes which occur in a familiar environment to the eyewitness. It is possible that CR techniques are not helpful in aiding memory recall in these circumstances because a familiar environment would automatically activate memories related to events other than the face, and so participants may not benefit from the environment as a retrieval cue (see Hupbachet al., 2011). This would be in line with the cue-overload principle (Watkins & Watkins, 1975), which indicates that the likelihood of recalling a target memory declines with an increasing number of retrieval cues. Future work could compare the effectiveness of Detailed CR for face construction following face encoding in both a familiar and an unfamiliar environment, with the expectation that Detailed CR only aids face construction if the environment is unfamiliar, but not familiar. This may also be dependent on the strength of the familiarity, and it may be valuable to subsequently investigate whether Detailed CR may still be effective in an environment of low familiarity, such as one that has only been visited once or twice prior to the incidence.

Effectiveness of Detailed CR may also be mediated by the age of the constructors. The thesis included participant-witnesses who were exclusively young to middle-aged adults (18 to 65 years), and it would be worthwhile investigating context effects for both children and older adults (aged over 65 years). Past research suggested that CR techniques are effective in eliciting a greater amount of recall for children of various ages (e.g., Dietze et al., 2010, 2013; Drohan-Jennings et al., 2010; Wagstaff et al., 2011), but no evidence seems to be available specifically in reference to face recognition or construction. With regard to older adults, inconsistent evidence has emerged. Some research has demonstrated context cues to be ineffective in improving face recognition (Memon et al., 2002; Searcy et al., 2001), whilst others have found that it increases performance to that of young adults (Wilcock et al., 2007). Therefore, it would be important to investigate how Detailed CR impacts on facial composite construction when constructors are either children (< 18 years) or older adults (> 65 years). Findings would inform forensic practitioners as to whether it would be appropriate and/or beneficial to employ this interviewing technique with witnesses of these ages.

In addition, it may also be worth considering other individual differences of witness-constructors. For example, individuals with Autism Spectrum Disorder (ASD) generally do not benefit from context cues derived through an MCR interview (Maras & Bowler, 2012), and so it is unlikely that Detailed CR would be effective for witnesses having this condition. Conversely, Physical CR appears to be more successful for these individuals (Maras & Bowler, 2012), and so it may be worthwhile investigating whether a Physical CR could be used for eyewitnesses with ASD to aid composite construction.

On a final note, the current research utilised a typical holistic system (EvoFIT) and a typical feature system (PRO-fit), and it would be sensible to attempt to replicate findings using other modern composite systems, such as EFIT-6 (a holistic system) as well as Identikit 2000 and FACES 4.0 (both feature systems used in the US). It is likely that Detailed CR effects would be reproduced given the similarities in procedures used with modern systems. Yet, it would be prudent to explore this empirically prior to informing forensic practitioners who use those alternative systems. It is also perhaps worth mentioning here that research is currently investigating the Detailed CR effect using Sketch composites, and initial data appears to be in favour of this technique also being effective for composite Sketches (Kuivaniemi-Smith, Richardson, Nash, Uther, & Frowd, 2018).

Conclusion

To summarise, the current thesis is the first to show a reliable benefit of utilising context cues during forensic face construction using modern systems. A Detailed CR, which promotes recall of the environmental context, was effective in facilitating facial composite construction via a typical holistic system (EvoFIT) and a typical feature system (PRO-fit), leading to higher correct naming and likeness ratings. The Detailed CR technique was effective to a greater extent as well as more consistently effective when participant-witnesses had reasonably good encoding of the environment in which the target face had been seen. Further, Detailed CR was also effective in combination with the H-CI, a technique which promotes character judgments of the face.

Both Detailed CR and H-CI are straightforward procedures and take little time to implement, thereby appearing to be valuable strategies to employ with eyewitnesses to increase identification of ensuing composites of offenders. Findings from the current thesis have already informed practice and Detailed CR consisting of the recall of the crime scene (without psychological state) is now regularly utilised by forensic practitioners in the UK and overseas when creating an EvoFIT composite (Frowd et al., 2019). It seems likely that results would generalise to other feature and recognition systems due to procedural similarities across these production systems.

REFERENCES

ACPO (2009). *Facial Identification Guidance*. National Policing Improvement Agency. Unpublished document.

Ackerman, B.P. (1985). Constraints on retrieval search for episodic information in children and adults. *Journal of Experimental Child Psychology*, *40*(1), 152-180.

Alison, L., Doran, B., Long, M.L., Power, N., & Humphrey, A. (2013). The effects of subjective time pressure and individual differences on hypotheses generation and action prioritization in police investigations. *Journal of Experimental Psychology: Applied*, *19*(1), 83-93.

Anderson, J.R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6, 451–474.

Anderson, J.A., & Bower, G.H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79(2)*, 97-123.

Aschermann, E., Dannenberg, U., & Schulz, A.P. (1998). Photographs as retrieval cues for children. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 12*(1), 55-66.

Ball, L.J., Shoker, J., & Miles, J.N. (2010). Odour-based context reinstatement effects with indirect measures of memory: The curious case of rosemary. *British Journal of Psychology*, *101*(4), 655-678.

Balota, D.A., Dolan, P.O., & Duchek, J.M. (2000). Memory changes in healthy older adults. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford handbook of memory*, (pp. 395-409). Oxford University Press: USA.

Barak, O., Vakil, E., & Levy, D.A. (2013). Environmental context effects on episodic memory are dependent on retrieval mode and modulated by neuropsychological status. *The Quarterly Journal of Experimental Psychology*, *66*(10), 2008-2022.

Beales, S.A., & Parkin, A.J. (1984). Context and facial memory: the influence of different processing strategies. *Human Learning*, *3(4)*, 257-264.

Bekerian, D.A., & Bowers, J.M. (1983). Eyewitness testimony: Were we misled?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 139-145.

Benson, P.J., & Perrett, D.I. (1991). Perception and recognition of photographic quality caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, *3*(1), 105-135.

Bowen, C.J., & Howie, P.M. (2002). Context and cue cards in young children's testimony: A comparison of brief narrative elaboration and context reinstatement. *Journal of Applied Psychology*, *87*(6), 1077-1085.

Bower, G.H. (1981). Mood and memory. *American Psychologist*, *36(2)*, 129-148.

Bower, G.H., & Cohen, P.R. (1982). Emotional influences in memory and thinking: Data and theory. In M.S. Clark & S.T. Fiske (Eds.), *Affect and cognition* (pp. 263-289). Hillsdale, NJ: Erlbaum.

Bower, G.H., & Karling, M.B. (1974). Depth of processing, pictures of faces and recognition of memory. *Journal of Experimental Psychology*, *103(4)*, 751-757.

Bower, G.H., Monteiro, K.P., & Gilligan, S.G. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior*, *17*(5), 573-585.

Bradlow, A.R., Nygaard, L.C., & Pisoni, D.B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & psychophysics*, *61*(2), 206-219.

Brainerd, C.J., Reyna, V.F., Howe, M.L., Kingma, J., & Guttentag, R.E. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development*, *55*(3-4, Serial No. 222), 109.

Brigham, J.C., Maass, A., Snyder, L.D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42*(4), 673-681.

Brinegar, K.A., Lehman, M., & Malmberg, K.J. (2013). Improving memory after environmental context change: A strategy of "preinstatement". *Psychonomic Bulletin & Review*, *20*(3), 528-533.

Brown, J.M. (2003). Eyewitness memory for arousing events: Putting things into context. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17*(1), 93-106.

Brown, C., Portch, E., Skelton, F.C., Fodarella, C., Kuivaniemi-Smith, H., Herold, K., Hancock, P.J.B., & Frowd, C.D. (2018). The impact of external facial features on the construction of facial composites. *Ergonomics, 62*(4), 575-592. DOI: 10.1080/00140139.2018.1556816

Bruce, V. (1979). Searching for politicians: An information processing approach to face recognition. *Quarterley Journal of Experimental Psychology, 31*(3), 373-395.

Bruce, V. (1982). Changing Faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*(1), 105–116.

Bruce, V., Henderson, Z., Newman, C., & Burton, A.M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7*(3), 207-218.

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.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception.* New York: Oxford University Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley, CA: University of California Press.

Brutsche, J., Cissé, A., Deléglise, D., Finnet, A., Sonnet, P., & Tiberghien, G. (1981). Effets de contexte dans la reconnaissance de visages non familiers. *CPC: European Bulletin of Cognitive Psychology*, *1*, 85-90.

Burri, C. (1931). The influence of an audience upon recall. *Journal of Educational Psychology*, *22*(9), 683-690.

Burton, A.M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, *11*(5), 429-433.

Campeanu, S., Craik, F.I., & Alain, C. (2015). Speaker's voice as a memory cue. *International Journal of Psychophysiology*, *95*(2), 167-174.

Campeanu, S., Craik, F.I., Backer, K.C., & Alain, C. (2014). Voice reinstatement modulates neural indices of continuous word recognition. *Neuropsychologia*, *62*, 233-244.

Campos, L., & Alonso-Quecuty, M.L. (1999). The cognitive interview: Much more than simply "try again". *Psychology, Crime and Law*, *5*(1-2), 47-59.

Carr, H. (1917). Maze studies with the white rat. I. *Normal animals, Animal Behavior, 7*, 259–275.

Christiaansen, R.E., & Ochalek, K. (1983). Editing misleading information from memory: Evidence for the coexistence of original and postevent information. *Memory & Cognition*, *11*(5), 467-475.

Chu, S., Handley, V., & Cooper, S.R. (2003). Eliminating context-dependent forgetting: Changing contexts can be as effective as reinstating them. *The Psychological Record*, *53*(4), 549-559.

Clark, M.S., Milberg, S., & Ross, J. (1983). Arousal cues arousal-related material in memory: Implications for understanding effects of mood on memory. *Journal of Memory and Language*, *22*(6), 633-649.

Clarke, C., & Milne, R. (2001). *A national evaluation of the PEACE Investigative Interviewing Course* (p. 187). London: Home office.

Clifford, B.R., & George, R. (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime and Law, 2*(3), 231-248.

Clifford, B.R., & Gwyer, P. (1999). The effects of the cognitive interview and other methods of context reinstatement on identification. *Psychology, Crime and Law*, *5*(1-2), 61-80.

Collins, R., Lincoln, R., & Frank, M.G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, Psychology, and Law, 9(1),* 69-78.

Cousins, R., & Hanley, J.R. (1996). The Effect of Environmental Context on Recall and Category Clustering Scores Following Relational and Individual Item Processing: A Test of the Outshining Hypothesis. *Memory, 1996, 4*(1), 79-90.

Craik, F.I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology*, *26*(2), 274-284.

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). Improving the reliability of eyewitness identifications: putting context into context. *Journal of Psychology*, *72*(2), 629-637.

Cutler, B.L., Penrod, S.D., O'Rourke, T.E., & Martens, T.K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behaviour*, *1*(2), 113-134.

Dalton, P. (1993). The role of stimulus familiarity in context dependent recognition. *Memory and Cognition, 21*(2), 223-234.

Dando, C.J. (2013). Drawing to remember: External support of older adults' eyewitness performance. *PloS one*, *8*(7), e69937.

Dando, C.J., Wilcock, R., & Milne, R. (2008). The cognitive interview: Inexperienced police officers' perceptions of their witness interviewing behaviour. *Legal and Criminological Psychology, 13*(1), 59-70.

Dando, C.J., Wilcock, R., & Milne, R. (2009). The Cognitive Interview: The Efficacy of a Modified Mental Reinstatement of Context Procedure for Frontline Police Investigators. *Applied Cognitive Psychology, 23*(1), 138-147.

Dando, C., Wilcock, R., Milne, R., & Henry, L. (2009). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology*, *23*(5), 698-716.

Davies, G., & Milne, A. (1982). Recognizing Faces in and out of Context. *Current Psychological Research*, *2*(4), 235-246.

Davies, G., & Milne, A. (1985). Eyewitness Composite Construction. A Function of Mental or Physical Reinstatement of Context. *Criminal Justice and Behaviour*, *12*(2), 209-220.

Davies, G., & Oldman, H. (1999). The impact of character attribution on composite production: A real world effect? *Current Psychology*, *18*(1), 128-139.

Davies, G., van der Willik, P., & Morrison, L.J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, *85*(1), 119-124.

Davies, G.M., & Little, M. (1990). Drawing on memory: Exploring the expertise of a police artist. *Medical Science and the Law*, *30*(4), 345-354.

Davies, G.M., Shepherd, J.W., & Ellis, H.D. (1978). Remembering faces: acknowledging our limitations. *Journal of the Forensic Science Society*, *18*(1), 19-24.

Davies, S. (2009). The effects of stress levels on the construction of facial composites. *Diffusion*, *2*, http://atp.uclan.ac.uk/buddypress/diffusion/?p=1763.

Davies-Thompson, J., Johnston, S., Tashakkor, Y., Pancaroglu, R., & Barton, J.J. (2016). The relationship between visual word and face processing lateralization in the fusiform gyri: A cross-sectional study. *Brain research*, *1644*, 88-97.

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*(5), 634-639.

Davis, J.P., Simmons, S., Sulley, L., Solomon, C., & Gibson, S. (2015). An evaluation of post-production facial composite enhancement techniques. *Journal of Forensic Practice*, *17*(4), 307-318.

Davis, J.P., Sulley, L. Solomon, C., & Gibson, S. (2010). A comparison of individual and morphed facial composites created using different systems. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A. T. Arslan (Eds.), *2010 IEEE International Conference on Emerging Security Technologies* (pp. 56–60). Canterbury: IEEE Computer Society.

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*, *150*(1), 102-118.

Davis, J.P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23*(4), 482-505.

Davis, M.R., McMahon, M., & Greenwood, K.M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology*, *19*(1), 75-93.

Davis, R., Sutherland, N.S., & Judd, B.R. (1961). Information content in recognition and recall. *Journal of Experimental Psychology*, *61(5)*, 422-429.

Dietze, P.M., Powell, M.B., & Thomson, D.M. (2010). Mental reinstatement of context with child witnesses: does it matter whether context is reinstated 'out loud'?. *Psychology, Crime & Law*, *16*(5), 439-448.

Dietze, P.M., Powell, M.B., & Thomson, D.M. (2012). Examination of the effect of mental reinstatement of context across developmental level, retention interval and type of mnemonic instruction. *Psychiatry, psychology and law*, *19*(1), 89-103.

Dietze, P.M., Sharman, S.J., Powell, M.B., & Thomson, D.M. (2013). Does free recall moderate the effect of mental context reinstatement instructions on children's cued recall?. *Psychology, Crime & Law*, *19*(10), 881-891.

Dietze, P.M., & Thomson, D.M. (1993). Mental re-instatement of context: A technique for interviewing child witnesses. *Applied Cognitive Psychology, 7*(2), 97-108.

Dodson, C.S., Johnson, M.K., & Schooler, J.W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition, 25*(2), 129-139.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental psychology*, *38*(3), 438-445.

Doherty-Sneddon, G., & Phelps, F.G. (2005). Gaze aversion: A response to cognitive or social difficulty?. *Memory & cognition*, *33*(4), 727-733.

Drohan-Jennings, D.M., Roberts, K.P., & Powell, M.B. (2010). Mental context reinstatement increases resistance to false suggestions after children have experienced a repeated event. *Psychiatry, Psychology and Law*, *17*(4), 594-606.

Dulsky, S.G. (1935). The effect of a change of background on recall and relearning. *Journal of Experimental Psychology*, *18*(6), 725-740.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological review*, *66*(3), 183-201.

Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*(4), 764-770.

Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 443-455.

Eich, J.E. (1980). The cue dependent nature of state dependent retrieval. *Memory and Cognition*, *8*(2), 157-173.

Eich, J.E., Weingartner, H., Stillman, R.C., & Gillin, J.C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Memory and Language*, *14*(4), 408-417.

Ellis, H.D. (1986). Face recall: A psychological perspective. *Human Learning: Journal of Practical Research & Applications, 5*, 1-8.

Ellis, H.D. (1990). Developmental trends in face recognition. *The Psychologist, 3,* 114-119.

Ellis, H.D., Davies, G.M., & Shepherd, J.W. (1978). A critical examination of the PhotoFIT system for recalling faces. *Ergonomics, 21*(4), 297-307.

Ellis, H.D., & Deregowski, J.B. (1981). Within-race and between-race recognition of transformed and untransformed faces. *The American journal of psychology*, *94*(1), 27-35.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8(4),* 431-439.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1980). The eterioration of Verbal Descriptions of Faces Over Different Delay Intervals. *Journal of Police Science and Administration*, *8(2)*, 101-106.

Emmett, D., Clifford, B.R., & Gwyer, P. (2003). An investigation of the interaction between cognitive style and context reinstatement on memory performance of eyewitnesses. *Personality and Individual Differences, 34*(8), 1495–1508.

Emmett, D., Clifford, B.R., Young, K., Kandova, M., & Potton, A. (2006). The use of a computer presented virtual licence plate and mental context reinstatement to improve recall accuracy of both licence plate and related event details. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(9), 1195-1207.

Evans, J.R., Marcon, J.L., & Meissner, C.A. (2009). Cross-racial lineup identification: Assessing the potential benefits of context reinstatement. *Psychology, Crime & Law, 15*(1), 19-28.

Eysenck, M.W. (1976). Arousal, learning and memory. *Psychological Bulletin, 83*(3), 389-404.

Farnsworth, P.R. (1934). Examinations in familiar and unfamiliar surroundings. *Journal of Social Psychology, 5*(1), 128-129.

Fernández, A., & Alonso, M.A. (2001). The relative value of environmental context reinstatement in free recall. *Psicológica*, *22*(2), 253-266.

Fernández, A., & Glenberg, A.M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, *13*(4), 333-345.

Fisher, R.P., & Craik, F.I.M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 701-711.

Fisher, R.P., & Craik, F.I.M. (1980). The effects of elaboration on recognition memory. *Memory and Cognition, 8*(5), 400-404.

Fisher, R., & Geiselman, R. (1992). *Memory Enhancing Techniques for Investigative Interviewing*. Springfield, IL: Charles Thomas Publishers.

Fisher, R.P., Geiselman, R.E., & Raymond, D.S. (1987). Critical analysis of police interview techniques. *Journal of Police Science and Administration*, *15*(3), 177-185.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, *83*(3), 323-336.

Fodarella, C., Brown, C., Lewis, A., & Frowd, C.D. (2015). Cross-age effects on forensic face construction. *Frontiers in psychology*, *6*, 1237.

Fodarella, C., & Frowd, C.D. (2013, September). Accuracy of Relational and Featural Information in Facial-Composite Images. In *2013 Fourth International Conference on Emerging Security Technologies* (pp. 16-20). IEEE.

Fodarella, C., Frowd, C.D., Warwick, K., Hepton, G., Stone, K., 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.

Fodarella, C., Kuivaniemi-Smith, H.J., Gawrylowicz, J., & Frowd, C.D. (2015). Detailed procedures for forensic face construction. *Journal of Forensic Practice, 17(4),* 271-280.

Frowd, C.D. (2010). Varieties of biometric facial techniques for detecting offenders. *Procedia Computer Science*, *2*(1), 3-10.

Frowd, C.D. (2011). Eyewitnesses and the use and application of cognitive theory. In G. Davey (Ed.) *Introduction to Applied Psychology* (pp. 267-289), BPS Wiley-Blackwell.

Frowd, C.D. (2012). Facial Recall and Computer Composites. In C. Wilkinson and C. Rynn (Eds). *Facial Identification* (pp. 42-56). Cambridge University Press: New York.

Frowd, C.D. (2017). Facial composite systems: Production of an identifiable face. In: *Face Processing: Systems, Disorders and Cultural Differences.* Nova Science Publishers.

Frowd, C.D., Bruce, V., McIntyre, A., & Hancock, P.J.B. (2007). The relative importance of external and internal features of facial composites. *British Journal of Psychology, 98(1),* 6-77.

Frowd, C.D., Bruce, V., Plenderleith, Y., & Hancock, P.J.B. (2006). Improving target identification using pairs of composite faces constructed by the same person. *IEE Conference on Crime and Security* (pp. 386-395), London: IET.

Frowd, C., Bruce, V., Ross, D., McIntyre, A., & Hancock, P.J. (2007). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition*, *15*(8), 954-984.

Frowd, C.D., Bruce, V., Smith, A., & Hancock, P.J.B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, *14(3)*, 276-287.

Frowd, C.D., Bruce, V., Storås, K., Spick, P. & Hancock, P.J.B. (2006). An evaluation of morphed composites constructed in a criminal investigation. In *Proceedings of the 16th Conference of the European Association of Psychology and Law* (pp. 59-66). London: IP-PA Publishing.

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L. McLanaghan, S. & Hancock, P.J.B. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, *11*(1), 33-52.

Frowd, C.D., Erickson, W.B., Lampinen, J.M., Skelton, F.C., McIntyre, A.H., & Hancock, P.J.B. (2015). A Decade of Evolving Composite Techniques: Regression- and Meta-Analysis. *Journal of Forensic Practice, 17*(4)*,* 319-334.

Frowd, C.D., & Fields, S. (2011). Verbalization effects in facial composite production. *Psychology, Crime & Law*, *17*(8), 731-744.

Frowd, C.D., Hancock, P.J.B., Bruce, V., McIntyre, A.H., Pitchford, M., Atkins, R., Webster, A., Pollard, J., Hunt, B., Price, E., & Morgan, S. (2010, September). Giving crime the'evo': catching criminals using EvoFIT facial composites. In *2010 International Conference on Emerging Security Technologies* (pp. 36-43). IEEE.

Frowd, C.D., Hancock, P.J.B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on applied perception (TAP)*, *1*(1), 19-39.

Frowd, C.D., Jones, S., Fodarella, C., Skelton, F.C., Fields, S., Williams, A., Marsh, J., Thorley, R., Nelson, L., Greenwood, L., Date, L., Kearley, K., McIntyre, A., & Hancock, P.J.B. (2014). Configural and featural information in facial-composite images. *Science and Justice, 54*(3), 215-227.

Frowd, C.D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C.G., & Hancock, P.J.B. (2007). An evaluation of US systems for facial composite production. *Ergonomics*, *50*(12), 1987-1998.

Frowd, C.D., Nelson, L., Skelton F.C., Noyce, R., Atkins, R., Heard, P., Morgan, D., Fields, S., Henry, J., McIntyre, A., & Hancock, P.J.B. (2012). Interviewing techniques for Darwinian facial composite systems. *Applied Cognitive Psychology*, *26*(4), 576-584.

Frowd, C.D., Pitchford, M., Skelton, F., Petkovic, A., Prosser, C., & Coates, B. (2012, September). Catching even more offenders with EvoFIT facial composites. In *2012 Third International Conference on Emerging Security Technologies* (pp. 20-26). IEEE.

Frowd, C.D., Portch, E., Killeen, A., Mullen, L., Martin, A.J., & Hancock, P.J.B. (2019). EvoFIT facial composite images: a detailed assessment of impact on forensic practitioners, police investigators, victims, witnesses, offenders and the media. In A. Stoica, G. Howells, K. McDonald-Maier, A. Erdogan, and T. Arslan (Eds.) *IEEE Proceedings of 2019 Eighth International Conference on Emerging Security Technologies*, July, University of Essex.

Frowd, C.D., Skelton, F., Atherton, C., Pitchford, M., Bruce, V., Atkins, R., Gannon, C., Ross, D., Young, F., Nelson, L. & Hepton, G. (2012). Understanding the multi-frame caricature advantage for recognizing facial composites. *Visual Cognition*, *20*(10), 1215-1241.

Frowd, C.D., Skelton, F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., McIntyre, A.H., & Hancock, P.J.B. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, *18*(2), 224-238.

Frowd, C.D., Skelton, F., Butt, N., Hassan, A., & Fields, S. (2011). Familiarity effects in the construction of facial-composite images using modern software systems. *Ergonomics*, *54*(12), 1147-1158.

Frowd, C.D., Skelton F.C., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C., & Hancock, P.J.B. (2013). Whole-face procedures for recovering facial images from memory. *Science and Justice*, *53(2)*, 89-97.

Gawrylowicz, J., Gabbert, F., Carson, D., Lindsay, W.R., & Hancock, P.J.B. (2012). Holistic versus featural facial composite systems for people with mild intellectual disabilities. *Applied Cognitive Psychology*, *26*(5), 716-720.

Geiselman, R.E., & Bjork, R.A. (1980). Primary versus secondary rehearsal in imagined voices: Differential effects on recognition. *Cognitive Psychology*, *12*(2), 188-205.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P., & Holland, H.L. (1985). Eyewitness memory enhancement in the police interview: cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70*(2), 401-412.

Gibling, F., & Davies, G. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology*, *79*(1), 129-141.

Glenberg, A.M. (1997). What memory is for. *Behavioral & Brain Sciences, 20*, 1-55.

Glenberg, A.M., Schroeder, J.L., & Robertson, D.A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & cognition*, *26*(4), 651-658.

Godden, D.R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: Land and underwater. *British Journal of Psychology*, *66*(3), 325-331.

Godden, D.R., & Baddeley, A.D. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*(1), 99-104.

Goldstein, A.G., & Mackenberg, E.J. (1966). Recognition of human faces from isolated facial features: A developmental study. *Psychonomic Science*, *6*(4), 149-150.

Green, D.L., & Geiselman, R.E. (1989). Building composite facial images: Effects of feature saliency and delay of construction. *Journal of Applied Psychology*, *74*(5), 714-721.

Greenspoon, J., & Ranyard, R. (1957). Stimulus conditions and retroactive inhibition. *Journal of Experimental Psychology*, *53*(1), 55-59.

Gruppuso, V., Lindsay, D.S., & Masson, M.E.J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review*, *14*(6), 1085-1089.

Hammond, L., Wagstaff, G.F., & Cole, J. (2006). Facilitating Eyewitness Memory in Adults and Children with Context Reinstatement and Focused Meditation. *Journal of Investigative Psychology and Offender Profiling*, *3(2)*, 117-130.

Hancock, P.J.B., Burke, K., & Frowd, C.D. (2011). Testing facial composite construction under witness stress. *International Journal of Bio-Science and Bio-Technology*, *3*(3), 65-71.

Hanczakowski, M., Zawadzka, K., & Coote, L. (2014). Context reinstatement in recognition: Memory and beyond. *Journal of Memory and Language*, *72*, 85-97.

Hanczakowski, M., Zawadzka, K., & Macken, B. (2015). Continued effects of context reinstatement in recognition. *Memory & Cognition*, *43*(5), 788-797.

Hasel, L.E., & Wells, G.L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and human Behavior*, *31*(2), 193-207.

Hayes, B.K., & Delamothe, K. (1997). Cognitive interviewing procedures and suggestibility in children's recall. *Journal of Applied Psychology*, *82*(4), 562-577.

Hershkowitz, I., Orbach, Y., Lamb, M.E., Sternberg, K.J., & Horowitz, D. (2002). A comparison of mental and physical context reinstatement in forensic interviews with alleged victims of sexual abuse. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(4), 429-441.

Hershkowitz, I., Orbach, Y., Lamb, M.E., Sternberg, K.J., & Horowitz, D. (2001). The effects of mental context reinstatement on children's accounts of sexual abuse. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(3), 235-248.

Hills, P.J., & Lewis, M.B. (2011). Reducing the own-race bias in face recognition by attentional shift using fixation crosses preceding the lower half of a face. *Visual Cognition*, *19*(3), 313-339.

Hills, P.J., & Lewis, M.B. (2005). Removing the own-race bias in face recognition by shifting attention. *Perception*, *34*(S), 206.

Hockley, W.E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1412-1429.

Hockley, W.E., Bancroft, T.D., & Bryant, E. (2012). Associative and familiarity-based effects of environmental context on memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *66*(2), 81-89.

Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, *23*(1), 65-74.

Hole, G.J., George, P.A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221-1240.

Hupbach, A., Gomez, R., & Nadel, L. (2011). Episodic memory updating: The role of context familiarity. *Psychonomic bulletin & review*, *18*(4), 787-797.

Itoh, Y. (2005). The facilitating effect of verbalization on the recogni- tion memory of incidentally learned faces. *Applied Cognitive Psychology, 19(*4), 421-433.

Jacoby, L.L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*(1), 21-38.

Jenkins, F., & Davies, G. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70*(1), 164-176.

Jensen, L.C., Harris, K., & Anderson, D.C. (1971). Retention following a change in ambient contextual stimuli for six age groups. *Developmental Psychology*, *4*(3), 394-399.

Johnston, R.A., & Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, *7*(2), 145-159.

Karpicke, J.D., Lehman, M., & Aue, W.R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of learning and motivation* (Vol. 61, pp. 237-284). Academic Press.

Kebbell, M.R., Preece, D., & Wagstaff, G.F. (1995). The effect of testimony elicited with a cognitive interview on jurors judgements of guilt and innocence. *Proceedings of the British Psychological Society, 4*(1), 59-67.

Kebbell, M.R., & Wagstaff, G.F (1999). The effectiveness of the cognitive interview. In D. Canter and L. Alison (Eds.), *Interviewing and deception* (pp. 23-41). Dartmouth: Ashgate.

Kehn, A., Renken, M.D., Gray, J.M., & Nunez, N.L. (2014). Developmental trends in the process of constructing own-and other-race facial composites. *The Journal of psychology*, *148*(3), 287-304.

Kerr, N.H., & Winograd, E. (1982). Effects of contextual elaboration on face recognition. *Memory & Cognition, 10*(6), 603-609.

Koen, J.D., Aly, M., Wang, W.C., & Yonelinas, A.P. (2013). Examining the causes of memory strength variability: Recollection, attention failure, or encoding variability?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1726-1741.

Koehn, C.E., & Fisher, R.P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. *Psychology, Crime and Law*, *3*(3), 209-218.

Kotchoubey, B., & El-Khoury, S. (2014). Event-related potentials indicate context effect in reading ambiguous words. *Brain and cognition*, *92*, 48-60.

Kovera, M.B., Penrod, S.D., Pappas, C., & Thill, D.L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, *82*(2), 235-246.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, *49*(1), 58-69.

Kuivaniemi-Smith, H.J., Nash, R.A., Brodie, E.R., Mahoney, G., & Rynn, C. (2014). Producing facial composite sketches in remote Cognitive Interviews: A preliminary investigation. *Psychology, Crime & Law*, *20*(4), 389-406.

Kuivaniemi-Smith, H.J., Richardson, B., Nash, R., Uther, M., & Frowd, C.D. (2018). The impact of mental reinstatement of context on sketch composites. European Association of Psychology and Law (EAPL), Turku, Finland, June.

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the Effectiveness of Pixelation and Blurring on Masking the Identity of Familiar Faces. *Applied Cognitive Psychology*, *15*(1), 101-116.

Laughery, K.R., Duval, G.C., & Fowler, R.H. (1977). *An analysis of procedures for generating facial images. Mug File Project Report Number UHMUG-2*. University of Houston, Texas.

Laughery, K.R., Duval, C., & Wogalter, M.S. (1986). Dynamics of facial recall. In A.W. Young (Eds.) *Aspects of face processing* (pp. 373-387). Dordrecht: Springer Netherlands.

Laughery, K.R., & Fowler, R.H. (1980). Sketch artist and Identikit procedures for recalling faces. *Journal of Applied Psychology, 65*(3), 307-316.

Lavie, N., Hirst, A., De Fockert, J.W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, *11*(5), 379-385.

Lee, K.J., & Perrett, D. (1997). Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces. *Perception*, *26*(6), 733-752.

Levy, D.A., Rabinyan, E., & Vakil, E. (2008). Forgotten but not gone: Context effects on recognition do not require explicit memory for context. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1620-1628.

Lewis, M.B. (2004). Face-Space-R: Towards a unified account of face recognition. *Visual Cognition, 11*(1), 29-69.

Lewis, M.B., & Johnston, R.A. (1998). Understanding caricatures of faces. *Quarterly Journal of Experimental Psychology, 51*(2), 321-346.

Lewis, M.B., & Johnston, R.A. (1999). A unified account of the effect of caricaturing faces. *Visual Cognition, 6*(1), 1-42.

Lind, S.E., & Bowler, D.M. (2010). Episodic memory and episodic future thinking in adults with autism. *Journal of Abnormal Psychology, 119*(4), 896-905.

Loftus, E.F. (1979). *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

Macken, W.J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 153–161.

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(2), 135-149.

Malpass, R.S., & Devine, P.G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*(4), 482-489.

Malpass, R.S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of personality and social psychology*, *13*(4), 330-334.

Mantyla, M. (1997). Recollections of faces: remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(5), 1203-1216.

Maras, K.L., & Bowler, D.M. (2012). Context reinstatement effects on eyewitness memory in autism spectrum disorder. *British Journal of Psychology*, *103*(3), 330-342.

Marsh, J. E., Patel, K., Labonté, K., Threadgold, E., Skelton, F. C., Fodarella, C., Thorley, R., Battersby, K.L., Frowd, C.D., Ball, L.J., & Vachon, F. (2017). Chatting in the face of the eyewitness: The impact of extraneous cell-phone conversation on memory for a perpetrator. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(3), 183-190.

Martin, A.J., Hancock, P.J.B., & Frowd, C.D. (2017, September). Breathe, relax and remember: An investigation into how focused breathing can improve identification of EvoFIT facial composites. In *2017 Seventh International Conference on Emerging Security Technologies (EST)* (pp.79-84). IEEE.

Martin, A.J., Hancock, P.J.B., Frowd, C.D., Heard, P., Gaskin, E., Ford, C., & Hewitt, T. (2018). EvoFIT composite face construction via practitioner interviewing and a witness-administered protocol. In G. Howells et al. (Eds.) *Proceedings of 12th NASA / ESA Conference on Adaptive Hardware and Systems*, 6th - 9th August, University of Edinburgh, Edinburgh, UK.

Matzel, L.D., Schachtman, T.R., & Miller, R.R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning & Motivation, 16*(4), 398-412.

Mayes, A., Meudell, P., & Neary, D. (1980). Do amnesics adopt inefficient encoding strategies with faces and random shapes? *Neuropsychologia*, *18*(4-5), 527-540.

McCauley, M.R., & Fisher, R.P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology*, *80*(4), 510-516.

McIntyre, A., Frowd, C.D., Bruce, V., & Hancock, P.J.B. (2010). Looking at people who are wearing glasses: the impact of glasses on matching photographs of unfamiliar faces and recognizing familiar composite faces. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A.T. Arslan (Eds.) *2010 IEEE International Conference on Emerging Security Technologies* (pp. 36-43).

McIntyre, A.H., Hancock, P.J.B., Frowd, C.D., & Langton, S.R. (2016). Holistic face processing can inhibit recognition of forensic facial composites. *Law and human behavior*, *40*(2), 128-135.

McQuiston-Surrett, D., Topp, L.D., & Malpass, R.S. (2006). Use of facial composite systems in U.S. law enforcement agencies. *Psychology, Crime and Law, 12*(5), 505-517.

Megreya, A.M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception, 38*(12), 1831-1848.

Meissner, C.A., & Brigham, J.C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3-35.

Meissner, C.A., Brigham, J.C., & Butz, D.A. (2005). Memory for own-and other-race faces: A dual-process approach. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(5), 545-567.

Memon, A., & Bruce, V. (1983). The effects of encoding strategies and context change on face recognition? *Human Learning*, *2*(4), 313-326.

Memon, A., & Bruce, V. (1995). Context effects in episodic studies of verbal and facial memory: A review. *Current Psychological Research and Reviews*, *4*(4), 349-369.

Memon, A., Holley, A., Wark, L., Bull, R., & Koehnken, G. (1996). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology*, *10*(6), 503-518.

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & cognition*, *30*(8), 1219-1227.

Memon, A., Meissner, C.A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*(4), 340-372.

Metzger, R.L., Boschee, P.F., Haugen, T., & Schnobrich, B.L. (1979). The classroom as learning context: Changing rooms affects performance. *Journal of Educational Psychology*, *71*(4), 440-442.

Meyer, D.E., & Schvaneveldt, R.W. (1976). Meaning, memory structures and mental processes. *Science, 192*(4234)*,* 27-33.

Milne, R., & Bull, R. (2002). Back to basics: A componential analysis of the original cognitive interview mnemonics with three age groups. *Applied Cognitive Psychology, 16*(7), 743-753.

Murnane, K., & Phelps, M.P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 882-894.

Murnane, K., Phelps, M.P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403-415.

Naka, M. (2012). Effect of interview techniques on children's eyewitness reports and subsequent memories of a viewed event. *Shinrigaku kenkyu: The Japanese journal of psychology*, *83*(4), 303-313.

Naveh-Benjamin, M., & Craik, F.I. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and aging*, *10*(2), 284-293.

Ness, H., & Bruce, V. (2006). An investigation into the use of CCTV footage to improve likeness in facial composites. In: *European Association of Psychology and Law Conference (EAPL 2006)*, 27-30 Jun 2006, Liverpool University, Liverpool.

Ness, H., Bruce, V. & Hancock, P.J.B. (2004). Combining facial composites and CCTV. Paper presented at the 14th European Conference on *Psychology and Law*, Krakow, July 2004.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, *30*(6), 755-764.

Odinot, G., Wolters, G., & van Giezen, A. (2012). Repeated suggestive questioning, consistency and the accuracy-confidence relation in eyewitness event memory. *Psychology, Crime and Law*, *18*(1), 27-47.

Parkin, A.J., & Hayward, C. (1983). The influence of trait and physical‒feature‒based orienting strategies on aspects of facial memory. *British Journal of Psychology*, *74*(1), 71-82.

Patterson, K.E., & Baddeley, A.D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(4), 406-417.

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are pre-schoolers sensitive to configural information in faces? *Developmental Science, 9(3)*, 270-277.

Penrod, S.D., Loftus, E.F., & Winkler, J. (1982). The reliability of eyewitness testimony: A psychological perspective. In N. Kerr & R. Bray (Eds.), *The psychology of the courtroom* (pp. 119-168). New York: Academic Press.

Penry, J. (1971). *Looking at Faces and Remembering Them.* London: Elek Books.

Perfect, T.J., Wagstaff, G.F., Moore, D., Andrews, B., Cleveland, V., Newcombe, S., Brisbane, K.A., & Brown, L. (2008). How can we help witnesses to remember more? It's an (eyes) open and shut case. *Law and Human Behavior*, *32*(4), 314-324.

Phelps, M.P. (1998). Mental reinstatement of environmental context in recognition memory. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 59*(6-B), 3085.

Rainis, N. (2001). Semantic contexts and face recognition. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(2), 173-186.

Read, J.D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology, 9*(2), 91-121.

Ready, D.J., Bothwell, R.K., & Brigham, J.C. (1997). The effects of hypnosis, context reinstatement, and anxiety on eyewitness memory. *International Journal of Clinical and Experimental Hypnosis*, *45*(1), 55-68.

Reder, L.M., Victoria, L.W., Manelis, A., Oates, J.M., Dutcher, J.M., Bates, J.T., Cook, S., Aizenstein, H.A., Quinlan, J., & Gyulai, F. (2013). Face Memory: Implications for theories of binding items to context. *Psychological science*, *24*(3), 363-372.

Reed, H.J. (1931). The influence of a change of conditions upon the amount recalled. *Journal of Experimental Psychology*, *14*(6), 632-649.

Reichler, G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*(2)*,* 275-280.

Rhodes, G., Brennan, S.E., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*(4), 473-494.

Richler, J.J., Cheung, O.S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464-471.

Russo, R., Ward, G., Geurts, H., & Scheres, A. (1999). When unfamiliarity matters: Changing environmental context between study and test affects recognition memory for unfamiliar stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 488-499.

Saufley, W.H., Otaka, S.R., & Bavaresco, J.L. (1985). Context effects: Classroom tests and contextual independence. *Memory & Cognition*, *13*(6), 522-528.

Schacter, D.L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 39-54.

Schreiber Compo, N., Hyman Gregory, A., & Fisher, R. (2012). Interviewing behaviors in police investigators: A field study of a current US sample. *Psychology, Crime & Law*, *18*(4), 359-375.

Schooler, J.W., & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*(1), 36-71.

Searcy, J.H., Bartlett, J.C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology*, *86*(2), 207-214.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*(2), 139-156.

Shaw, J.S., Bjork, R.A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic bulletin & review*, *2*(2), 249-253.

Shaw, J.S., & McClure, K.A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, *20*(6), 629-653.

Shepherd, J.W., & Ellis, H.D. (1996). Face recall-methods and problems. *Psychological issues in eyewitness identification*, 87-115.

Shrimpton, S., Oates, K., & Hayes, S. (1998). Children's memory of events: effects of stress, age, time delay and location of the interview. *Applied Cognitive Psychology*, *12*(2), 133-143.

Skelton, F.C., Frowd, C.D., & Speers, K.E. (2015). The benefit of context for facial-composite construction. *Journal of Forensic Practice, 17*(4)*,* 281-290.

Smith, S.M. (2013). Effects of environmental context on human memory. In *The SAGE handbook of applied memory*, *2*, 162-182.

Smith, S.M. (1985). Environmental context and recognition memory reconsidered. *Bulletin of the Psychonomic Society, 23*(3), 173-176.

Smith, S.M. (1988). Environmental context-dependent memory. In G.M. Davies and D.M. Thomson (Eds), *Memory in context: Context in memory* (pp. 13-34). New York: Wiley.

Smith, S.M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning & Memory*, *5*(5), 460-471.

Smith, S.M. (1994). Theoretical principles of context-dependent memory. *Theoretical aspects of memory*, 2, 168-195.

Smith, S.M., Glenberg, A.M., & Bjork, R.A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342-353.

Smith, S.M., & Sinha, A.K. (1987). Effects of brief immersion in a flotation tank on memory and cognition (Tech. Rep. No. CSCS-004). College Station, TX: Texas A&M University, Committee for the Study of Cognitive Science.

Smith, S.M., & Vela, E. (1986). Outshining-the relative effectiveness of cues. *Bulletin of the Psychonomic Society, 24*(5), 350-350.

Smith, S.M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology*, *6*(2), 125-139.

Smith, S.M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203-220.

Smith, S.M., & Manzano, I. (2010). Video contextdependent recall. *Behavior Research Methods*, *42*(1), 292–301.

Solomon, C.J., Gibson, S.J., & Maylin, M. (2012). EFIT-V: Evolutionary algorithms and computer composites. In C. Wilkinson, & C. Rynn (Eds.), *Craniofacial identification* (pp. 24-41). Cambridge: Cambridge University Press.

Sporer, S.L. (1993). Clothing as a contextual cue in facial recognition. *German Journal of Psychology, 17*, 183-199.

Sporer, S.L. (2007). Person descriptions as retrieval cues: Do they really help?. *Psychology, Crime & Law*, *13*(6), 591-609.

Steinberg, S. (2012). Facial identification catalogs, Available at samanthastein-berg.-com/new_facial_identification_catalog.html.

Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225-245.

Tanaka, J.W., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory & cognition*, *25*(5), 583-592.

Thomson, D.M. (1981). Person identification: Influencing the outcome. *Australian & New Zealand Journal of Criminology*, *14*(1), 49-54.

Thomson, D.M., Robertson, S.L., & Vogt, R. (1982). Person recognition: The effect of context. *Human Learning*, *1*(1), 37-54.

Tibon, R., Vakil, E., Goldstein, A., & Levy, D.A. (2012). Unitization and temporality in associative memory: Evidence from modulation of context effects. *Journal of Memory and Language*, *67*(1), 93-105.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology-Psychologie Canadienne, 26*(1)*,* 1-12.

Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology, 66*(4)*,* 319-327.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory, *Psychological* Review, *80*(5), 359-380.

Tulving, E., & Thomson, D.M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology, 87*(1)*,* 116-124.

Tulving, E., & Watkins, M.J. (1973). Continuity between recall and recognition. *American Journal of Psychology*, *86*(4), 739-748.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology - Human Experimental Psychology, 43(2)*, 161-204.

Valentine, T. (2001). Face-space models of face recognition. *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*, 83-113.

Valentine, T., Davis, J.P., Thorner, K., Solomon, C., & Gibson, S. (2010). Evolving and combining facial composites: Between-witness and within-witness morphs compared. *Journal of Experimental Psychology: Applied*, *16*(1), 72-86.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing - The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology - Human Experimental Psychology*, *44*(A), 671-703.

Wagstaff, G.F. (1982). *Context effects in eyewitness reports*. Paper presented at the Law and Psychology Conference. Swansea, Wales, July.

Wagstaff, G.F., Brunas-Wagstaff, J., Cole, J. & Wheatcroft, J. (2004). New directions in forensic hypnosis: facilitating memory with focused meditation. *Contemporary Hypnosis*, *21*(1), 14-27.

Wagstaff, G.F., Brunas-Wagstaff, J., Knapton, L., Winterbottom, J., Crean, V., Cole, J. & Wheatcroft, J. (2004). Facilitating memory with hypnosis, focused meditation and eyeclosure. *International Journal of Clinical and Experimental Hypnosi*s, *52*(4), 434-455.

Wagstaff, G.F., Cole, J., Wheatcroft, J., Marshall, M., & Barsby, I. (2007). A componential approach to hypnotic memory facilitation: Focused meditation, context reinstatement and eye movements. *Contemporary Hypnosis*, *24*(3), 97-108.

Wagstaff, G.F., Wheatcroft, J.M., Caddick, A.M., Kirby, L.J., & Lamont, E. (2011). Enhancing witness memory with techniques derived from hypnotic investigative interviewing: Focused meditation, eye-closure, and context reinstatement. *International Journal of Clinical and Experimental Hypnosis*, *59*(2), 146-164.

Walker-Smith, G.J. (1978). The effects of delay and exposure duration in a face recognition task. *Perception & Psychophysics*, *24*(1), 63-70.

Walker-Smith, G.J. (1980). Memorizing facial identity, expression and orientation. *British Journal of Psychology, 71*(3), 415-424.

Watkins, M.J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behaviour*, *15*(5), 505-517.

Watkins, O.C., & Watkins, M.J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1*(4)*,* 442-452.

Wells, G.L., Ferguson, T.J., & Lindsay, R.C. (1981). The tractability of eyewitness confidence and its implications for triers of fact. *Journal of Applied Psychology*, *66*(6), 688-696.

Wells, G.L., Lindsay, R.C., & Ferguson, T.J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology*, *64*(4), 440-448.

Wells, G.L., Memon, A., & Penrod, S.D. (2007). Eyewitness evidence: improving its probative value. *Psychological Sciences in the Public Interest*, *7*(2), 45-75.

Wheeler, D.D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59-85.

Wilcock, R.A., Bull, R., & Vrij, A. (2007). Are old witnesses always poorer witnesses? Identification accuracy, context reinstatement, own-age bias. *Psychology, Crime & Law*, *13*(3), 305-316.

Wilhite, S.C. (1991). Evidence of a negative environmental reinstatement effect. *British Journal of Psychology, 82*(3)*,* 325-342.

Winograd, E., & Rivers-Bulkeley, N.T. (1977). Effects on changing context on remembering faces. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(4), 397-405.

Wixted, J., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025-1054.

Wong, C.K., & Read, J.D. (2011). Positive and negative effects of physical context reinstatement on eyewitness recall and identification. *Applied Cognitive Psychology*, *25*(1), 2-11.

Wright, A.M., & Alison, L. (2004). Questioning sequences in Canadian police interviews: Constructing and confirming the course of events?. *Psychology, Crime and Law*, *10*(2), 137-154.

Yamada, K., Nabeta, T., Oka, K., & Chujo, K. (2009). Effect of environmental context on false recognition. *Shinrigaku Kenkyu: The Japanese Journal of Psychology, 80*(2), 90-97.

Yarmey, A.D. (1979). *The psychology of eyewitness testimony* (pp. 204-05). New York: Free Press.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747-759.

SUPPLEMENTARY EXPERIMENT: Combining Context Reinstatement and Holistic-Cognitive Interview

Stage 1: Composite construction

Design

A 2 (System: EvoFIT; PRO-fit) × 2 (CR: Minimal, Detailed) × 2 (Interview: CI, H-CI) between-participants design was utilised. Following the unfamiliar face encoding in an unfamiliar environment (café), participant-witnesses underwent Minimal or Detailed CR after a 24-hour-delay, were interviewed via CI or H-CI and created a single EvoFIT or PRO-fit composite.

Participants

Sixty-four (26 males, 38 females; M_{age} = 28.0, SD_{age} = 9.6 years) UCLan staff and students volunteered who were recruited on the basis of being unfamiliar with the target identities.

Materials

Target faces were eight photographs of current Rugby League players (Chris Annakin, Jacob Fairbank, Jamil Fakir, Darrell Goulding, Ben Harrison, Lance Hohaia, Joel Monaghan and Ian Thornley). They were sourced from the Internet and were in good quality, showing a full frontal face with a minimal facial expression. Stimuli were printed in colour (8 cm × 10 cm).

Procedure

Each participant viewed a target face in an unfamiliar café on the University campus, and met with the researcher the following day in a different room to construct a composite of this face. The procedure was the same as in Experiment 2.

Stage 2: Composite naming

Design, Materials and Procedure

The same three-factor design was used as described in Stage 1. Materials were 64 composites and eight target photographs, printed as before, and named in the same way as in Experiment 1 and 2.

Participants

Sixty-four (62 males, 2 females; M_{age} = 44.0; SD_{age} = 15.1 years) participants volunteered. They were recruited as part of an opportunity sample on the UCLan campus on the basis of being familiar with the target identities.

Stage 3: Composite rating

Design, Materials and Procedure

The Design was the same as in Stage 1, but within-participants. Materials were the same as in Stage 1, and Procedure was the same as in Experiment 1 and 2.

Participants

Eighteen (7 males, 11 females; M_{age} = 38.7, SD_{age} = 16.0 years) UCLan staff and students volunteered, unfamiliar with the target identities.

Results

Composite naming

As before, participant responses to both targets and composites were scored for accuracy. The target pictures were named again very accurately, at 98.8% correct. Correct composite naming rates are shown in Table 1.

Naming scores were analysed using Independent Samples ANOVA for CR (Minimal, Detailed), System (EvoFIT, PRO-fit) and Interview (CI, H-CI). There was a non-significant main effect of CR [$F(1,56) = 0.14$, $p = .71$, $\eta_p^2 < .01$], but a significant main effect of Interview [$F(1,56) = 4.57$, $p = .037$, $\eta_p^2 = .08$], with H-CI leading to higher-named composites than CI. There was also a significant main effect of System [$F(1,56) = 41.10$, $p < .001$, $\eta_p^2 = .42$], with EvoFIT composites being named higher than PRO-fit composites. The three two-way interactions and the three-way interaction were not significant ($p > .26$, $F < 1.29$).

Table 1. Percentage of EvoFIT and PRO-fit composites correctly named by Context Reinstatement and Interview type

	Interview†				
	CI		H-CI		
System†	Minimal	Detailed	Minimal	Detailed	Mean
EvoFIT	23.4 (17.0)	20.3 (14.8)	34.4 (18.6)	29.7 (13.3)	27.0 (16.2)
PRO-fit	4.7 (6.5)	6.3 (6.7)	7.8 (6.5)	9.4 (8.8)	7.0 (7.1)
Mean	13.7 (14.3)		20.3 (17.0)		17.0 (16.0)

Note: † Significant main effect of Interview, $p < .05$; * Significant main effect of System, $p < .001$.

Composite rating

The pattern of means for likeness ratings (Table 2) again reflect those for naming. Repeated Measures ANOVA revealed non-significant main effects of both CR [$F(1,17) = 1.97$, $p = .18$, $\eta_p^2 = .10$] and System [$F(1,17) = 2.24$, $p = .15$, $\eta_p^2 = .12$]. However, there was a significant main effect of Interview [$F(1,17) = 36.34$, $p < .001$, $\eta_p^2 = .68$] as composites produced following an H-CI received higher likeness ratings than those following a CI.

There was also one significant interaction, between System and CR [$F(1,17) = 5.35$, $p = .033$, $\eta_p^2 = .24$] (other interactions, $p > .25$, $F < 1.41$). EvoFIT composites were actually rated significantly *worse* following Detailed CR compared to Minimal CR [$t(17) = 3.51$, $p = .003$], with no reliable difference for PRO-fit composites ($p = .55$). Also, following Minimal CR, EvoFIT composites were rated significantly better than PRO-fit composites [$t(17) = 2.35$, $p = .031$]; there was no reliable difference between EvoFIT and PRO-fit composites constructed after Detailed CR ($p = .60$).

Table 2. EvoFIT and PRO-fit mean composite rating (SD) by Context Reinstatement and Interview type

	Interview†				
	CI		H-CI		
System	Minimal	Detailed	Minimal	Detailed	Mean
EvoFIT	3.1 (0.7)	2.8 (0.7)	3.5 (0.9)	3.2 (0.8)	3.1 (0.8)
PRO-fit	2.6 (0.7)	2.6 (0.5)	3.1 (0.8)	3.2 (0.7)	2.8 (0.7)
Mean	2.8 (0.7)		3.2 (0.8)		3.0 (0.8)

Note: † Significant main effect of interview, $p < .001$; Significant CR × System interaction, $p < .05$. The rating scale was from 1 (poor likeness) to 7 (good likeness).

Discussion

Detailed CR was ineffective (cf. Minimal CR) at improving naming or likeness ratings. Whilst means for naming and rating of PRO-fit composites were in the expected direction, albeit non-significant, likeness ratings for EvoFIT significantly *decreased* for composites produced following Detailed CR (cf. Minimal CR). For interview type, composites produced following H-CI were, as hypothesised, significantly better named and rated than those produced following CI. Composites were named reliably better when constructed from EvoFIT than PRO-fit.