Memory

Face value: testing the utility of contextual face cues for face recognition

Nina Tupper, James D. Sauer, Melanie Sauerland, Isabel Fu & Lorraine Hope

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

Face value: testing the utility of contextual face cues for face recognition

Nina Tupper[a,b], James D. Sauer[b,c], Melanie Sauerland[a], Isabel Fu[a] and Lorraine Hope[b]

[a]Clinical Psychological Sciences, Maastricht University, Maastricht, the Netherlands; [b]Department of Psychology, University of Portsmouth, Portsmouth, UK; [c]Department of Psychology, University of Tasmania, Hobart, TAS, Australia

ABSTRACT

The presence of multiple faces during a crime may provide a naturally-occurring contextual cue to support eyewitness recognition for those faces later. Across two experiments, we sought to investigate mechanisms underlying previously-reported cued recognition effects, and to determine whether such effects extended to encoding conditions involving more than two faces. Participants studied sets of individual faces, pairs of faces, or groups of four faces. At test, participants in the single-face condition were tested only on those individual faces without cues. Participants in the two and four-face conditions were tested using no cues, correct cues (a face previously studied with the target test face), or incorrect cues (a never-before-seen face). In Experiment 2, associative encoding was promoted by a rating task. Neither hit rates nor false-alarm rates were significantly affected by cue type or face encoding condition in Experiment 1, but cuing of any kind (correct or incorrect) in Experiment 2 appeared to provide a protective buffer to reduce false-alarm rates through a less liberal response bias. Results provide some evidence that cued recognition techniques could be useful to reduce false recognition, but only when associative encoding is strong.

The recognition and identification of an unfamiliar face is a difficult task, and becomes even more challenging as the number of unfamiliar faces to be remembered increases. Indeed, accuracy rates for face recognition and identification decrease when there are multiple faces to study (Clifford & Hollin, 1981). Although divided attention at encoding may play a role in this effect (e.g., Bindemann, Mike Burton, & Jenkins, 2005; Bindemann, Jenkins, & Burton, 2007), this does not entirely account for the multiple-face recognition disadvantage. Accuracy decreases even when participants have unlimited time to encode the presented faces and when divided attention is controlled for (Bindemann, Sandford, Gillatt, Avetisyan, & Megreya, 2012). The persistence of the multiple-face disadvantage suggests that the drop in accuracy is caused, at least in part, by an increase in memorial demand when attempting to hold two faces in memory for any period of time (Bindemann et al., 2012).

Interestingly, the presence of multiple faces not only increases memorial demand, but also provides a naturally-occurring contextual cue that may promote recognition. Early experiments demonstrated increased recognition rates for a previously-studied face when that face, having been paired with another face at study, was presented with its corresponding pair at test (rather than when presented alone or with a different face; e.g.,

Watkins, Ho, & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977). It is possible that the benefits of cued recognition could be used to support eyewitness memory in the context of multiple perpetrator crimes, whereby the face of the first perpetrator might enhance the eyewitness' ability to accurately identify the second perpetrator in a lineup. However, experiments attempting to apply cuing in the context of eyewitness identification for multiple perpetrator crimes do not align with these theoretical predictions. In one experiment, for example, eyewitnesses saw a two-perpetrator crime and subsequently viewed two sequential lineups presented side-by-side: Eyewitnesses viewed one picture from the lineup of the perpetrator next to one picture from the lineup of the accomplice, and this was repeated until all of the photos of each lineup had been shown. Compared with the traditional sequential and simultaneous lineups, this adapted cuing lineup improved accuracy in rejecting innocent suspects from a perpetrator-absent lineup (i.e., decreased false alarms), but not in identifying actual perpetrators (i.e., increasing hits; Wells & Pozzulo, 2006). In another experiment, eyewitnesses of a two-perpetrator crime viewed a simultaneous lineup of the perpetrator with a single photo of the accomplice presented next to it as a cue, or vise-versa (Dempsey, 2012). Compared with when no cue was presented, this technique improved the correct

rejection of innocent suspects for the perpetrator, but not the correct rejection of innocent suspects for the accomplice.

It is possible that results of these adapted lineups did not demonstrate the benefits of cued face recognition because of a mismatch in converting this theoretically informed technique to practice. These results may also be due to the original experiments not having been designed with an eyewitness context in mind, therefore failing to control for potentially confounding variables that exist in such a context. To address these limitations, it would be useful to re-examine what we understand about cued face recognition with an eye towards its eventual application in the context of eyewitness identification for multiple perpetrator crimes. Therefore, we sought to replicate previously-reported enhancing effects in cued face recognition, to determine whether such effects could extend to include more than two faces (as many crimes involve more than two perpetrators), and to investigate the mechanisms underlying those effects. We compared the traditional cued encoding condition in which participants study paired faces, with two additional encoding conditions: a control condition where participants encoded single faces and another experimental condition where participants encoded groups of four faces.

Cued recognition and faces

Memory researchers have long known that context matters for retrieval. Contextual cues are often implemented to help individuals recall seemingly-forgotten details in episodic memory (encoding specificity principle; Thomson & Tulving, 1970), including to facilitate eyewitness recall during investigative interviews (context reinstatement; Geiselman, Fisher, MacKinnon, & Holland, 1986). It has been hypothesised that these cues work because the retrieval of a memory is dependent upon the way it was stored, and an item in episodic memory is, by nature, nested within our experience of the relevant event (Polyn, Norman, & Kahana, 2009; Tulving & Thomson, 1973). Thus, episodic memory is not only tied to temporal markers (when an event occurred), but can also be integrated with memory traces for the context of the event. For example, we often recall items in semantic clusters (i.e., related words like book, hardcover, paperback, bestseller) as a result of our long-standing associations between items in our experience, but also in source clusters as a result of the associations we form during study phase between items (i.e., information from a friend vs. the internet; Polyn et al., 2009). A contextual cue at retrieval takes advantage of the associative nature of memory whereby, for example, peripheral details of the environment can cue additional pathways for the retrieval of critical details.

Contextual cues also benefit recognition memory, including recognition for faces (e.g., Watkins et al., 1976). A variety of external contexts have been shown to enhance recognition for target faces, including backgrounds on which the faces were studied, eyewitness descriptions of targets, clothing worn by targets, and other faces presented with the target faces (see Davies, 1988 for a review). For example, Winograd and Rivers-Bulkeley (1977) asked participants to memorise pairs of faces during the study phase—one of which served as the target face while the other served as the cue. During the test phase, participants were presented with a target face alongside either a face they had previously studied (*correct cue*), a face that had not been previously studied (*incorrect cue*), or no cue at all. In this forced-choice paradigm, recognition performance for target faces was enhanced by the presentation of correct cue faces and impaired by incorrect cue faces, while performance with no cues fell in-between. Similarly, Watkins et al. (1976) demonstrated reduced hit rates for face recognition when they implemented *substituted cues*: previously-studied faces that had been paired with a different face during study. In other words, simply swapping context, as opposed to introducing new context, also affected face recognition performance.

Although the above experiments found an enhancing effect of using correct cues for face recognition, other research has found no such effect (e.g., Bower & Karlin, 1974) or found that hit rates were undermined by incorrect cues, but not enhanced by correct cues (Kan, Giovanello, Schnyer, Makris, & Verfaellie, 2007). A number of issues relating to encoding and recognition may underpin these discrepant results. First, the context must be strongly encoded in association with the target for recognition (Peris, 1985, as cited in Davies, 1988). If the participant did not pay attention to the contextual information, or did not link it with the target information, then presenting the cue at test will not improve recognition. Second, it appears that context is useful as a recognition memory cue only when other, stronger cues are lacking (Smith & Vela, 2001). Some theories of recognition hold that recognition is comprised of two mechanisms (e.g., Mandler, 1980). The first is the perceptual system that is automatically activated and produces fast answers that hinge on the feeling of familiarity. The second is the cognitive system, which is activated when the first does not immediately provide an answer. The second system is slower and searches for external information, like context, to aid the response. This process is also captured in the rationale of the *outshining hypothesis,* which contends that we use the most relevant cues available to recognise faces. When our memory trace is strong, that memory *outshines* the utility of environmental context. However, for weak memory traces, such as when there was suboptimal encoding or longer retention intervals, context may support memory to improve performance (Smith & Vela, 2001).

The current research

Cued recognition presents an interesting means of enhancing face recognition, a concept that may prove useful to

the applied fields of eyewitness identification or wanted-persons recognition. The second face in a two-perpetrator crime provides a naturally-occurring context for the eyewitness, and one that is particularly relevant for humans. Given our natural tendency to orient attention towards other human faces, a second face may increase the chance of incidental associative encoding (Di Giorgio, Turati, Altoè, & Simion, 2012). Thus, face cues may support eyewitnesses of a multiple perpetrator crime while viewing a suspect lineup. This is not an entirely novel idea; at least two published experiments (Hobson & Wilcock, 2011; Wells & Pozzulo, 2006) and one unpublished dissertation (Dempsey, 2012) have attempted to apply face cuing to support eyewitness memory and adapt identification procedures in the context of multiple-perpetrator crimes. However, none of these attempts provided convincing evidence that this method of cuing memory could aid lineup identification decisions, and it is of interest to understand why this might be. The current research aims to understand the disconnect between a theoretically intriguing mnemonic device and the poorly-understood difficulties in applying such a device; particularly by extending it to explore boundary conditions and orienting the effects. These goals are explored further below.

To begin, previous experiments showing inconsistent effects of face cues expose concerns for the replicability of a cued face recognition effect. Relevant experiments were mostly conducted in the 1970s, and it is only decades later that we are now interested in using this research for applied contexts (i.e., eyewitness identification). However, original studies made decisions that are incongruent with contemporary methodological standards in psychology. For one, the relevant research reports methodologies with confounding variables. For example, Winograd and Rivers-Bulkeley (1977) specifically paired male-female pairs according to overt compatibility, with particular care to maintain similar ages between them. Further, participants rated perceived compatibility of the couples, further enhancing the romantic link between the pairs. It is possible that this likeness between the pairings provided contextual information for correct and incorrect cuing. A participant, for example, might correctly recognise a target face of a confident-looking 20-year-old white male paired with a confident-looking 20-year-old white female because the pairing enhances recognition. It might also be because the nature of the pairing gives the participant a hint as to the correct answer given prior knowledge that the two faces presented are an intuitively compatible couple. By extension, a participant might be worse at recognising that same male when paired with an older, shy-looking female because the participant intuits they are not a compatible couple. Likewise, previous experiments using face pairs did not randomise the left-right orientation of those pairs such that if the target face was on the right side of the screen during the encoding phase, it was also presented on the right side during test (Watkins et al., 1976; Winograd & Rivers-Bulkeley, 1977). It

may be that the boost in hit rates is sensitive to this spatial context as well as contextual cuing. The recent movement in psychological science to replicate previous effects stems in part from a realisation that the field now has updated knowledge on methodological issues like sample sizes, randomisation, and experimenter influence that have changed the way we conduct experiments (i.e., Simmons, Nelson, & Simonsohn, 2011). Given the intriguing theoretical rationale for the effects, but inconsistent results obtained, we aimed to replicate the cued recognition effect isolated from other influences such as the context effects of placement, or intuitive responses based on overt compatibility of couples. Therefore, we randomised left-right location of target faces during encoding and randomly paired or grouped photographs of male faces selected from a large database.

Furthermore, we sought to place the cued face recognition effects in context by comparing them to traditional recognition memory, whereby single faces are encoded and single faces are tested; for simplicity, we refer to this as *orienting* the effects of cuing because we aim to understand how memory tested in cued recognition paradigm compares to memory tested in a traditional, non-cued recognition paradigm. In the cued face recognition paradigm, researchers use correct cues to reconstruct encoding context, incorrect cues to represent a change in context, and no cues as control trials. However, not presenting a cue at test is also a change of context because the absence of context is itself a deviation from the original context. It is therefore unclear whether previously-reported effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both. To accommodate this, we included a new control condition in which participants encoded two or four faces in the encoding phase, but were not presented with cues during test phase. Figure 1 shows a graphic representation of encoding and test conditions. This control condition (panel A) allowed us to examine the impact of number of faces at encoding (Figure 1, panel B, C). It may be that recognition rates without cues are equivalent across single-face and multiple-face encoding conditions. Thus, any benefit of correct cuing, for example, would be an enhancement over general face recognition. By contrast, if the no-cue condition varies between the single-face and multiple-face conditions, then any benefit of correct cuing would be compensation for increased difficulties.

Existing research also exposes questions regarding the potential boundary conditions for such an effect. For example, do beneficial context effects vary according to the number of additional faces to be encoded? Research on cued face recognition to date has held encoding conditions constant while manipulating the conditions at retrieval (e.g., Watkins et al., 1976; Winograd & Rivers-Bulkeley, 1977). However, pairing two faces at encoding represents the minimum number of faces individuals might encode when attempting to implement cued face recognition. Limiting our consideration to pairs of faces fails to reflect the variability in group sizes that individuals
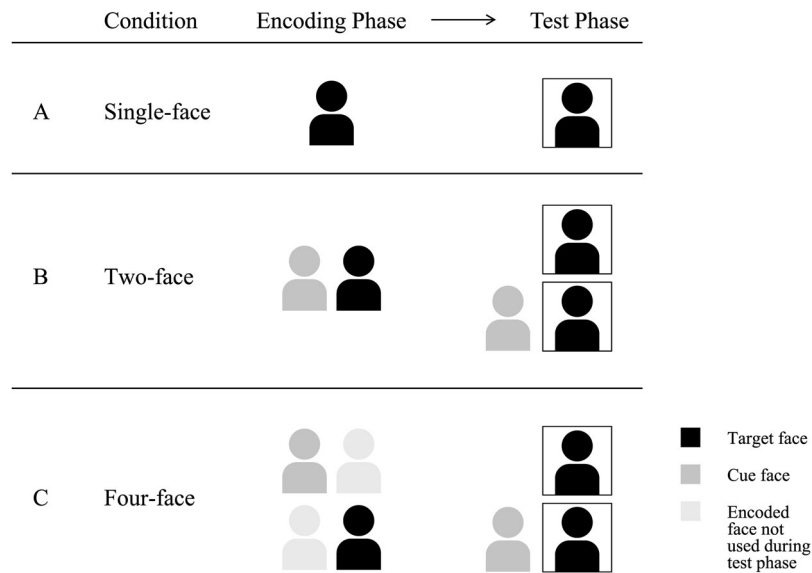
Figure 1. Experiments 1 and 2: Procedure for encoding and testing. Participants studied single faces, pairs of faces, or groups of four faces during the encoding phase. Participants in the single-face condition were only tested on individual faces (i.e., no-cue trials). Participants in the two-face and four-face condition were tested on trials with no cue and trials with either a correct cue or incorrect cue. Here, we demonstrate in conditions when, during the test phase, both the target face is present and the cue face is correct. Color coding is used to demonstrate how cues and targets were chosen, but pariticipants did not know which faces during the encoding phase would be used as targets or cues. Targets and cues were determined at random from the left or right position of the encoding phase, but were always presented on the right and left, respectively, during the test phase.

encounter every day, and it is unclear if the benefits identified in previous work extend to conditions in which participants need to encode more stimuli. In other words, if four faces are studied together, can *one* of those faces effectively serve as a cue for another?

There are several reasons why cuing with four-face stimuli may not be as effective as with two-face stimuli. First, there is the potential for generalised deficits in memory with the increased cognitive load associated with having more faces (e.g., four vs. two) to encode. Cuing in this case may be helpful, but with a general trend for the four-face condition to have decreased accuracy. Or cuing could be less effective because there is diminished memory to support. Second, when faces are always presented at study as pairs, a correct cue also represents a complete cue. However, if four faces are presented during encoding and only one face is used to cue the recognition of the target face, this may represent a partial cue. Therefore, providing only one of the set of the associated cues at test may show a weaker effect. Third, there is the potential that associations between multiple (i.e., more than two) faces may not be encoded equally, and therefore may not be equally effective at retrieval. While two faces viewed side-by-side provide only one association that can be formed at encoding, four faces provide multiple associations that can be formed with varying strengths. Thus, the face chosen by experimenters to be a contextual cue may not be the one that was most strongly associated with the target at encoding. By contrast, it may be that each face is one of several possible cues, meaning that any of the faces would be sufficient to enhance recognition.

Experiments 1 and 2

These applied and theoretical considerations converge into three questions for the current research: (1) do previously-reported effects of contextual cuing replicate with contemporary methodology? (2) how do cuing effects vary as a function of the number of cues to be encoded? and (3) do these effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both?

Experiment 1 was a first attempt to establish to what extent correct cuing is beneficial or compensating for increased cognitive demand, and also whether this effect endures when cognitive load is increased by viewing multiple (i.e., more than two) faces at once. However, because results yielded minimal effects of either face encoding condition or cue type, we added a manipulation to more strongly encourage associative encoding between faces in Experiment 2. Thus, Experiment 2 was a direct replication of Experiment 1, with the addition of the manipulation during the encoding phase (see procedure for further details). In both Experiments 1 and 2, we expected to replicate previous effects of cued face recognition in the two-face condition, and that those effects would extend to the four-face condition, though weaker.

Method

Participants

A total of 159 participants were recruited from Maastricht University. Three failed to follow instructions and were excluded from data analysis. The remaining 156

participants (121 women) in Experiments 1 ($n = 78$) and 2 ($n = 78$) were between the ages of 18 and 43 ($M = 22.14$, $SD = 3.09$). Participants were compensated with either participation credit or a 5€ (Experiment 1) or 7.50€ (Experiment 2) voucher. These studies were approved by the standing ethical committee of the Faculty of Psychology and Neuroscience at Maastricht University.

Design

We used a 3 (encoding condition: single face, two faces, four faces) × 3 (cue type: no cue, correct cue, incorrect cue) × 2 (target face presence: old, new) mixed design, with encoding condition as the between-subjects factor. Participants were randomly assigned to one of the three between-subjects conditions. Cue type was randomly ordered and target faces were determined at random. Two versions of each experiment were created in which the order of trials for all conditions was varied, and in the two- and four-face conditions, different target faces and cue faces were selected.

Materials

Face stimuli

For each experiment, 204 photographs of male faces were used from a database of faces at Flinders University (Adelaide, Australia), the Psychological Image Collection at Stirling (PICS, (n.d.); pics.stir.ac.uk), and the AR Face Database (Martinez & Benavente, 1998). Of those, 36 faces were used as target faces and the remaining 168 faces were used as correct cue faces at test, incorrect cue faces at test, or filler faces during encoding. Faces with features that were highly distinctive (e.g., piercings, unique haircuts, facial hair) were replaced. All photographs were in full colour with a resolution of 300 × 300 pixels. They were presented on a computer screen with a resolution of 1366 × 768 pixels using Microsoft PowerPoint 2010. In the single-face condition, faces were placed in the centre of the screen. In the two-face condition, two faces were placed side by side. In the four-face condition, the faces were presented in a 2 × 2 matrix. To create the multiple faces conditions, the faces were randomly grouped into

pairs (two-face condition) or groups of four faces (four-face condition). Images of faces were not repeated between encoding blocks. In other words, each encoding block presented the participant with a new set of target faces that they had never seen before. Target faces were the same photographs at both encoding and test. See Figure 2 for example stimuli for each of the three encoding conditions.

Methodological considerations

In comparing effects across conditions with only single faces and groups of four faces, there were two major methodological concerns: participant allocation of attention during encoding, and the number of total trials encoded. Each of these is addressed below.

Allocation of attention

We considered how increasing the number of faces on a given trial would impact the allocation of attention at encoding. Bindemann et al. (2012) demonstrated that given two stimuli, participants generally allocate their attention evenly between the stimuli. Unsurprisingly, identification performance was positively correlated with the amount of time spent studying the appropriate target face. Thus, the amount of attention devoted to any individual stimulus would be reduced if encoding duration remained constant, but the number of to-be-encoded stimuli would increase. We attempted to attenuate this concern by providing comparative encoding durations per face, rather than per trial. Participants in all conditions were allotted 2 s per face, even if those faces were presented in pairs or as a group. Thus, the exposure duration for single-face, two-face and four-face trials were 2, 4, and 8 s, respectively. Participants were instructed to focus on all faces, but whether participants actually allocated attention equally across stimuli was beyond our control.

Encoding load

In order to compare the single-, two- and four-face conditions, we needed participants to complete the same number of test trials. However, this presents a problem at
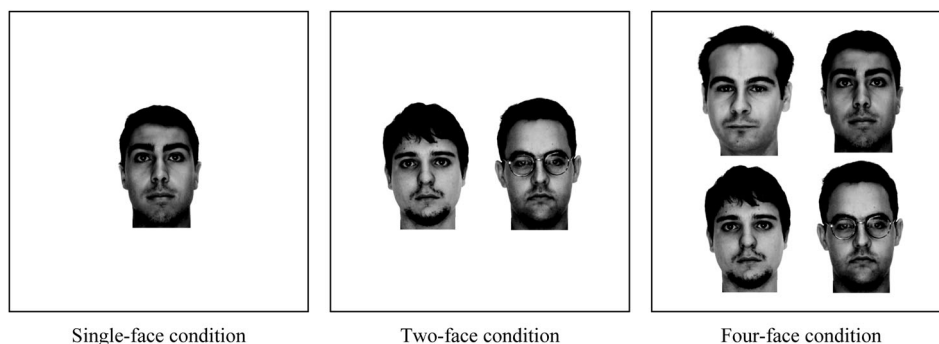

Figure 2. Experiments 1 and 2: Example encoding stimuli using images from AR Face database.

encoding. A participant who studies 36 trials of individual faces in the encoding phase will be tested on 72 trials of individual faces (half old, half new faces) in the test phase. Participants in the two-face and four-face conditions should therefore also confront 72 test trials with old and new faces. However, to test for cued face recognition, the 72 target faces at test require extra faces at encoding to be used as cues. The natural solution to this is to hold constant the number of encoding trials so that all conditions study 36 trials at encoding. Yet 36 trials in a two-face condition and a four-face condition consequently mean that a participant studies 72 faces and 144 faces at encoding, presenting extreme differences in memory load. In short, if we hold constant the number of test trials, we cannot also control for both the number of encoding trials and the total number of faces studied.

Our solution was to create testing blocks so that participants only ever studied 36 faces in each block: 36 individual faces, 18 paired faces, and 9 groups of four faces. They were then tested on 72, 36, and 18 old and new faces in the different conditions. The two- and four-face conditions repeated this study-test cycle until they had also completed 72 test trials, meaning the two-face condition encountered two study-test blocks and the four-face condition encountered four study-test blocks. Breaks in between the blocks were included in order to compensate for the memory load induced by viewing more faces.

Procedure

Experiment 1
Participants arrived in the lab to take part in individual sessions. Before the experiment, participants completed a practice tutorial with three encoding trials and three test trials. Faces presented during the tutorial were not used during the actual experiment.

At encoding, participants were asked to memorise a number of faces. They were instructed to pay attention to the faces and that each face would be displayed for 2 s (if in the one-face condition), or that each set of faces would be displayed for 4 s (if in the two-face condition), or 8 s (if in the four-face condition). Encoding trials advanced automatically after the given interval and a fixation cross appearing for 500 ms between trials. Between the encoding and testing phases, participants performed a short arithmetic task for 30 s. At test, each trial was presented for 5 s, after which time the programme automatically advanced to display a fixation cross. Participants were instructed to judge a test face, as either being old (previously-studied) or new (never-before-seen). If the participant indicated an "old" response, they were also asked to indicate their confidence on a scale from 1 (*not at all confident)* to 7 (*very confident).*[1] Participant responses were self-paced while the fixation cross was displayed, so that participants pressed the spacebar on the keyboard to advance to the next trial. Response latencies were not recorded.

In each condition, only one of the faces presented during the encoding phase was used as a target face at test. Participants in the single-face condition completed one study-test block, consisting of 36 single face stimuli during the encoding phase, and 72 single face trials at test (including the 36 study faces as targets and 36 previously unseen faces; these are referred to as old and new faces, respectively).

Single-face condition. During the encoding phase, participants viewed 36 individual photos of faces. Each face (all target faces) was shown for a duration of 2 s. As for all conditions, each trial was followed by an interstimulus interval (ISI) of 500 ms before the next trial appeared. During the testing phase, participants completed 72 single face trials, half of which were old faces and half of which were new.

Two-face condition. During the encoding phase, participants viewed 18 pairs of faces. Each pair (one target face and one cue face) was shown for 4 s (allowing 2 s per face). No information was provided to participants on which of the two faces they would be tested, but participants were told to focus on all faces. During the testing phase, participants completed 36 trials: 12 were presented with no cue, 12 with a correct cue (a previously-studied face), and 12 with an incorrect cue (a never-before-seen face). Presence of the target face was manipulated so that half of the trials were old, and half were new. The target face was randomly selected from the pairs of faces for each of the two versions of the test-phase. At test, the target face was always shown on the right side of the screen during test and always designated by a green square. This process was repeated with 18 new pairs of faces during encoding. Thus, there were two study-test blocks.

Four-face condition. This condition involved four blocks of trials. During the encoding phase, the participants viewed nine groups of four faces. Each stimulus group (one target face and three potential cue faces) was shown for 8 s (2 s per face). Participants were instructed to focus on all faces during encoding. During the testing phase, participants completed 18 trials: six trials were presented with no cue, six trials with a correct cue, and six trials with an incorrect cue. Presence of the target face was manipulated so that half of the trials were old, and half were new. The target face was randomly selected from the groups of faces and was always presented on the right side of the screen during test and designated by a green square. This process was completed three more times with nine new groups of four faces for each encoding phase.

Experiment 2
Experiment 2 was a direct replication of Experiment 1 except for the addition of a manipulation during the encoding phase to strengthen associative encoding.

Participants were told that the experiment was to test the supposed relationship between initial impressions of how criminal the defendant looks, the memorability of their faces, and the number of years a single convicted defendant or a group of convicted defendants were sentenced to prison. After each stimulus was presented during the encoding stage, participants were asked to rate on a Likert scale how many years (4–10) the person(s) presented had been sentenced to prison for by the judge. For example, in the four-face condition, participants were instructed, "*All members of the pair were sentenced to the same number of years in prison. If you mark the number "4", it means that the person on the top left was sentenced to 4 years in prison, and the person on the top right, the bottom left, and the bottom right were each sentenced to 4 years in prison.*" This manipulation led to a second difference between Experiments 1 and 2. In Experiment 1, the encoding trials and ISIs automatically advanced throughout the encoding phase. In Experiment 2, although the encoding trials automatically advanced after 2 s per face, the ISI was self-paced so that participants had unlimited time to their sentencing judgment and then pressed the spacebar on the keyboard to advance to the next stimulus.

Results

Preparation of data

Data were screened for outliers and normality prior to analysis. Shapiro-Wilk tests were used to assess normality for each of the conditions. In Experiment 2, one outlier was removed from the four-face condition because of an abnormally high false-alarm rate. When normality was violated, both nonparametric (Kruskal Wallis H and Mann-Whitney U) and standard inferential tests (one-way Analyses of Variance [ANOVAs] and t-tests) were conducted on hit rates and false-alarm rates. Results did not differ as a result of test, therefore ANOVAs and t-tests are reported throughout. Where relevant, we conducted Bayesian analyses – run in JASP (2017) software – to determine whether the data provided evidence of equivalence. We use Jarosz and Wiley's (2014) interpretations of Bayes Factors as evidence for the alternative hypothesis, in which they report descriptive thresholds provided by Jeffreys (1961) and Raftery (1995). Approximate cut-offs

are as follows: 1–3 constitutes anecdotal/weak evidence for the null or alternative hypothesis, 3–10 is positive/substantial evidence; 10–20 is strong/very strong evidence; 20+ is very strong/decisive.[2] Descriptive statistics of hit and false-alarm rates as a function of cue type and face group are reported in Table 1. Sensitivity and response bias rates as a function of cue type and face group are reported in Table 2. Inferential statistics are not reported in text, but are available in Tables 3–5.

Hit rates were defined as the probability of a correct response given that the stimulus (S) is present (i.e., the trial shows an old face) and were calculated by dividing the number hits on present trials by the total number of present trials. The false-alarm rate was the probability of an incorrect response given that the stimulus is absent (i.e., the trial shows a new face) and were calculated by dividing the number of false alarms on absent trials by the total number of absent trials. We computed signal detection statistics to test for effects on sensitivity and response bias (d' and c, respectively). In the current research, sensitivity is the capacity to discriminate between old and new faces, which can range from 0 to 4.65. A score of 0 indicates participants cannot discriminate at all, while 4.65 is considered an effective ceiling for discrimination (Macmillan & Creelman, 2005). Response bias is the tendency to respond *old* or *new* regardless of whether the face is actually old or new. Scores can range from −2.33 to +2.33. In our experiments, negative scores reflect the tendency to say "old" which is considered liberal response bias, and positive scores reflect the tendency to say "new", which is considered a conservative response bias (Macmillan & Creelman, 2005).

In order to examine recognition performance the no-cue condition, we tested hit rates and false-alarm rates across all face conditions (in single-, two-, four-face). Given that the single-face condition does not have correct and incorrect cues, only the two- and four-face groups were considered when comparing these cue type.

Replicating and extending effects of cued face recognition

First, we tested for the replication of previous effects of cued face recognition, and whether these effects vary as

Table 1. Experiments 1 and 2: mean hit rates and false-alarm rates (95% CI) by cue type and face encoding condition.

	Hit rates by cue type				False-alarm rates by cue type			
	Correct	Incorrect	None	Overall	Correct	Incorrect	None	Overall
No. faces encoded								
Experiment 1								
Single-face	–	–	.70 (.65–.74)	.70 (65–.74)	–	–	.19 (.15–.23)	.19 (.15–.23)
Two-face	.71 (.64–.78)	.72 (.64–.80)	.71 (.64–.78)	.71 (.66–.77)	.25 (.19–.30)	.23 (.16–.30)	.25 (.19–.32)	.24 (.19–.30)
Four-face	.72 (.67–.78)	.74 (.68–.80)	.76 (.71–.81)	.74 (.70–.78)	.14 (.10–.19)	.18 (.12–.23)	.26 (.21–.31)	.19 (.15–.24)
Experiment 2								
Single-face	–	–	.77 (.72–.82)	.77 (.72–.81)	–	–	.15 (.11–.18)	.15 (.11–.18)
Two-face	.72 (.66–.78)	.70 (.66–.74)	.71 (.66–.76)	.71 (.68–.74)	.19 (.14–.24)	.14 (.10–.18)	.23 (.16–.30)	.19 (.15–.23)
Four-face	.72 (.66–.79)	.73 (.64–.81)	.74 (.68–.80)	.73 (.67–.79)	.15 (.09–.20)	.21 (.16–.25)	.25 (.20–.31)	.20 (.17–.24)

Note: CI = Confidence Interval.

Table 2. Experiments 1 and 2: mean sensitivity and response bias rates (95% CI) by cue type and face encoding condition.

	Sensitivity by cue type (d')				Response bias by cue type (c)			
	Correct	Incorrect	None	Overall	Correct	Incorrect	None	Overall
No. faces encoded								
Experiment 1								
Single-face	–	–	1.47 (1.31–1.63)	1.47 (1.31–1.63)	–	–	0.19 (0.07–0.30)	0.19 (0.07–0.30)
Two-face	1.24 (0.98–1.49)	1.29 (1.00–1.58)	1.22 (0.90–1.55)	1.40 (1.20–1.60)	0.07 (–0.08–0.23)	0.06 (–0.16–0.28)	–0.02 (–0.10–0.07)	0.07 (–0.09–0.23)
Four-face	1.68 (1.36–2.00)	1.56 (1.30–1.82)	1.36 (1.11–1.61)	1.60 (1.37–1.83)	0.19 (0.08–0.29)	0.13 (–0.04–0.29)	–0.02 (–0.10–0.07)	0.10 (–0.01–0.21)
Experiment 2								
Single-face	–	–	1.94 (1.67–2.20)	1.94 (1.67–2.20)	–	–	0.17 (0.03–0.31)	0.17 (0.03–0.31)
Two-face	1.50 (1.26–1.74)	1.57 (1.39–1.76)	1.10 (0.85–1.35)	1.54 (1.36–1.71)	0.11 (–0.04–0.25)	0.24 (0.13–0.34)	0.03 (–0.13–0.19)	0.20 (0.09–0.30)
Four-face	1.61 (1.36–1.86)	1.44 (1.16–1.72)	1.26 (1.05–1.47)	1.54 (1.35–1.73)	0.23 (0.06–0.40)	0.09 (–0.05–0.23)	–0.02 (–0.17–.13)	0.14 (0.00–0.27)

Note: CI = Confidence Interval.

a function of the number of faces to be encoded. Therefore, we conducted 2 (encoding condition: two-face, four-face) × 3 (cue type: no, correct, incorrect) mixed ANOVAs on hit rates and false-alarm rates with cue type as the within-subjects variable (see Table 3 for inferential statistics). We expected that correct cues would enhance hit rates compared with no cues and incorrect cues. Although we also expected to see false-alarm rates rise with correct cues (vs. no cues or incorrect cues), we did not expect these to eliminate the enhancing effects of correct cuing for recognition.

Hit rates

In both Experiments 1 and 2, there were no significant main effects of face encoding condition or cue type on hit rates; the interaction was also non-significant. Thus, we found no evidence that either the number of faces at encoding or the type of cue presented at test affected participants' hit rates. Because we wanted to test evidence in favour of the null hypothesis, we conducted Bayes analyses using JASP (JASP Team, 2017). In these analyses, models are presented for (i) each main effect, (ii) the combined main effects, and (iii) the combined main effects with the interaction of the main effects. The strength of evidence for each model is then compared against the strength of evidence for the null hypothesis. Bayes analyses provided strong support for the null hypothesis in both experiments, with the model including both main effects and the interaction resulting in a Bayes factor (BF_{01}) of 233.55 or higher when compared with the evidence in favour of the null. In other words, none of the models (including individual main effects, combined main effects, or combined main effects and the interaction) fit the data better than the null hypothesis.

False-alarm rates

In Experiment 1, there was a significant main effect of cue type that was modified by a significant interaction between face encoding condition and cue type. Inspection of simple main effects showed that on correct cue trials, participants in the two-face condition had significantly more false alarms relative to the four-face condition. However, these effects should be interpreted cautiously. While we report the results of the interaction because of the significant p-value, Bayesian analyses provided weak support for the interaction ($BF_{10} = 2.81$). All other comparisons were non-significant.

In Experiment 2, the main effect of cue type was significant and Bayes analyses provided substantial evidence in support of a meaningful effect ($BF_{10} = 8.21$). Collapsing across encoding face condition, participants produced more false alarms on no-cue trials compared with correct- or incorrect-cue trials, though there was no significant difference between the incorrect-cue trials and the correct-cue trials. The main effect of face encoding condition was not significant and Bayes analyses provided evidence in favour of the null hypothesis ($BF_{01} = 3.75$).

Table 3. Experiments 1 and 2: results for ANOVAs and *t*-tests comparing hit rates and false-alarm rates across cue type and face encoding condition.

	df	F	η^2	t	d	p
Experiment 1						
Orient the effect of cuing						
One-way ANOVAs (No. faces; 1, 2, 4)						
HR	2, 75	1.69	.043			.191
FAR	2, 75	1.81	.046			.171
Replicate and extend cuing effects						
Mixed ANOVAs 2 (No. faces: 2v4) × 3 (cue type)						
HR						
Cue	2, 100	0.27	.005			.762
No. faces (2v4)	1, 50	0.61	.012			.439
Interaction	2, 100	0.41	.008			.666
FAR						
Cue	2, 100	4.74	.087			.011
No. faces (2v4)	1, 50	2.26	.043			.139
Interaction	2, 100	3.23	.061			.044
Follow-up simple main effects						
Independent sample *t*-tests (No. faces: 2v4)						
No cue	50			0.15	.04	.881
Correct cue	50			2.80	.78	.007
Incorrect cue	50			1.21	.33	.234
Experiment 2						
Orient the effect of cuing						
One-way ANOVAs (No. faces; 1, 2, 4)						
HR	2, 74	1.10	.029			.337
FAR	2, 74	3.95	.096			.023
Replicate and extend cuing effects						
Mixed ANOVAs 2 (No. faces: 2v4) × 3 (cue type)						
HR						
Cue	2, 98	0.10	.002			.904
No. faces (2v4)	1, 49	0.35	.007			.556
Interaction	2, 98	0.18	.004			.833
FAR						
Cue	1.78, 87.12	5.89	.107			.005
No. faces (2v4)	1, 49	0.32	.006			.577
Interaction	1.78, 87.12	2.96	.057			.063

Notes: *t*-tests are two-tailed. HR and FAR represent hit rates and false-alarm rates, respectively.

Sensitivity and response bias

We used *t*-tests to compare sensitivity and response bias across the two- and four-face conditions (see Table 5 for inferential statistics). In Experiments 1 and 2, there were no significant effects. In Experiment 1, Bayesian analyses only provided sufficient evidence to support the null hypothesis for response bias ($BF_{01} = 3.39$), but not sensitivity ($BF_{01} = 1.83$). In Experiment 2, Bayesian analyses only provided sufficient evidence to support the null hypothesis for sensitivity ($BF_{01} = 3.53$), but not response bias ($BF_{01} = 2.88$).

We used repeated measures ANOVAs to compare sensitivity and response bias across the no-, correct-, and incorrect-cue trials. In Experiment 1, there were no significant differences between groups for either sensitivity or response bias, and there was sufficient evidence to support the null hypothesis for sensitivity ($BF_{01} = 5.71$). In Experiment 2, however, response bias did change as a result of cue type ($BF_{10} = 4.26$). Participants had a negative response bias on no-cue trials (i.e., were more likely to be biased to respond *old*) compared with trials with correct and incorrect cues. In Experiment 2, sensitivity also changed as a result of cue type such that participants had lower sensitivity (i.e., were less able to discriminate old from new faces) on no-cue trials compared with correct or incorrect cue trials. However, Bayesian analysis provided only anecdotal evidence for this effect ($BF_{10} = 1.43$).

Orienting the effect of cuing

Next, we tested whether previously-reported effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both. To address this, we tested the effect of group size on the no-cue trials on hit rates and false-alarm rates, using one-way ANOVAs with three levels (face encoding condition: one, two, or four faces). This allows us to understand the initial differences in how number of faces at encoding impacts recognition performance.

Hit rates

In both Experiments 1 and 2, there were no significant differences in hit rates between face encoding conditions. This means that there was no evidence of any difference between the single-, two-, or four-face encoding conditions when no cue was presented during test. Bayesian analyses showed anecdotal support for the null hypothesis in Experiment 1 ($BF_{01} = 2.45$), but moderate support for the null hypothesis in Experiment 2 ($BF_{01} = 3.80$).

False-alarm rates

For Experiment 1, there were again no differences in false-alarm rates between groups. Bayesian analyses did not provide compelling evidence to support the null hypothesis ($BF_{01} = 2.24$). For Experiment 2, however, there was a

Table 4. Experiments 1 and 2: results for Bayesian ANOVAs comparing hit rates and false-alarm rates across cue type and face encoding condition.

Models	P(M)	P(M|data)	BF_M	BF_{10} (BF_{01})	% error
Experiment 1					
Orient the effects of cuing					
One-way ANOVAs (No. faces; 1, 2, 4)					
HR					
Null model	0.50	0.71	2.45	1.00	
No. faces	0.50	0.29	0.41	0.41 (2.45)	0.01
FAR					
Null model	0.50	0.69	2.24	1.00	
No. faces (1, 2, 4)	0.50	0.31	0.45	0.45 (2.24)	0.01
Replicate and extend effects of cuing					
Mixed ANOVAs 2 (No. faces: 2v4) × 3 (cue type)					
HR					
Null model	0.20	0.68	8.51	1.00	
Cue	0.20	0.06	0.23	0.08 (12.36)	0.80
No. faces (2v4)	0.20	0.24	1.27	0.35 (2.82)	1.27
Cue + 2v4	0.20	0.02	0.08	0.03 (32.99)	2.91
Cue + 2v4 + Cue*2v4	0.20	<0.01	0.01	.004 (233.55)	1.72
FAR					
Null model	0.20	0.11	0.48	1.00	
cue	0.20	0.30	1.72	2. 83 (0.35)	0.80
No. faces (2v4)	0.20	0.08	0.32	0.70 (1.42)	0.80
Cue + 2v4	0.20	0.22	1.13	2.07 (0.48)	2.35
Cue + 2v4 + Cue*2v4	0.20	0.30	1.68	2.81 (0.36)	2.12
Experiment 2					
Orient the effects of cuing					
One-way ANOVAs (No. faces; 1, 2, 4)					
HR					
Null model	0.50	0.79	3.80	1.00	
No. faces (1, 2, 4)	0.50	0.21	0.26	0.26 (3.80)	0.03
FAR					
Null model	0.50	0.31	0.44	1.00	
No. faces (1, 2, 4)	0.50	0.70	2.28	2.28 (0.44)	0.01
Replicate and extend effects of cuing					
Mixed ANOVAs 2 (No. faces: 2v4) × 3 (cue type)					
HR					
Null model	0.20	0.71	9.90	1.00	
Cue	0.20	0.05	0.21	0.07 (14.23)	0.83
No. faces (2v4)	0.20	0.22	1.12	0.31 (3.25)	1.02
Cue + 2v4	0.20	0.02	0.07	0.02 (43.48)	5.84
Cue + 2v4 + Cue*2v4	0.20	<0.01	<0.01	<0.01 (353.33)	2.62
FAR					
Null model	0.20	0.07	0.30	1.00	
Cue	0.20	0.58	5.50	8.21 (0.12)	0.96
No. faces (2v4)	0.20	0.02	0.08	0.27 (3.75)	1.14
Cue + 2v4	0.20	0.15	0.72	2.16 (0.46)	1.00
Cue + 2v4 + Cue*2v4	0.20	0.18	0.87	2.54 (0.39)	1.24

Notes: HR and FAR represent hit rates and false-alarm rates, respectively. BF_{10} refers to evidence in favor of the alternative hypothesis. In parentheses, BF_{01} refers to evidence in favor of the null hypothesis ($BF_{01} = 1/ BF_{10}$). BF_M refers to the change between prior (P[M] and posterior P[M|data] model odds (see Wagenmakers et al., 2017 for in-depth explanation of Bayesian hypothesis testing).

significant difference between face encoding conditions, though Bayesian analyses provided only weak or anecdotal support for this ($BF_{10} = 2.28$). The single-face condition had significantly fewer false alarms compared with the four-face condition. The false-alarm rate for the two-face condition fell between, and was neither significantly greater than the single-face nor significantly smaller than the four-face condition.

Summary of key findings

In Experiment 1, neither hit rates nor false-alarm rates were impacted by the type of cue provided at test (correct, incorrect, no cue). On no-cue trials, face encoding condition had a statistically significant but negligible effect on false-alarm rates, such that the four-face condition produced a higher false-alarm rate than the single-face condition. In Experiment 2, which included an additional manipulation designed to promote associative encoding, cuing again had no impact on hit rates, but reduced false alarms. This effect was present regardless of the face encoding condition.

Discussion

Previous research has shown that when faces are studied in pairs, one of those faces can be used at test to cue recognition of the other. Across two experiments, we sought to answer the following three questions in this cued face recognition domain: (1) do previously reported effects replicate with updated methodology? (2) how do cuing effects vary as a function of the number of cues to be

Table 5. Experiments 1 and 2: results for *t*-tests and ANOVAs comparing sensitivity (*d′*) and response bias (*c*) as a factor of face encoding condition or cue type.

	df	F	η^2	t	d	p	BF$_{10}$ (BF$_{01}$)
Experiment 1							
Two vs. four-face conditions							
d′	49			−1.28	−0.36	.207	0.55 (1.83)
c	43.97			−0.35	−0.10	.728	0.30 (3.39)
Cue type (no, correct, incorrect cue)							
d′	2, 60	0.74	.024			.482	0.18 (5.71)
c	1.60, 47.98	2.17	.067			.135	0.53 (1.87)
Experiment 2							
Two vs. four-face conditions							
d′	48			−0.06	−0.02	.949	0.28 (3.53)
c	48			0.71	0.20	.488	0.35 (2.88)
Cue type (no, correct, incorrect cue)							
d′	2, 54	3.23	.107			.047	1.43 (0.70)
c	2, 54	4.71	.149			.013	4.26 (0.24)

Notes: We computed signal detection statistics to test for effects on sensitivity and response bias (d′ and c, respectively). We used *t*-tests to compare across the two-face and four-face conditions. We used repeated measures ANOVA's across the no, correct, and incorrect cue conditions. The single-face condition was excluded due to the difference in cue types. BF$_{10}$ refers to evidence in favor of the alternative hypothesis. In parentheses, BF$_{01}$ refers to evidence in favor of the null hypothesis.

encoded? and (3) do these effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both? While relevant research has only used pairs of faces at encoding, we aimed to extend previous investigations of this effect by adding two conditions (single-face control, four-face condition) to the encoding phase. Specifically, the single-face condition was added to establish baseline differences in studying one vs. multiple faces, thus orienting the effects of cuing. The four-face condition was added to determine whether the effect would also arise in situations in which people encode more than two faces at a time. At test, participants in the single-face condition were shown individual faces that had either been previously studied (old) or had never been studied before (new). Participants in the two-face or four-face encoding conditions were shown either a target face alone (no cue), or a target face presented alongside a correct or incorrect cue face. Experiment 2 was a replication of Experiment 1, but with the addition of a rating task to strengthen the associative encoding of the multiple faces.

Experiment 1 provided little evidence with which to answer our research questions, and it is only in Experiment 2, when we added a manipulation to enhance associative encoding, that cued recognition effects arose. This alone suggests that cuing effects are likely to require more than the arbitrary pairing of stimuli at encoding. Accordingly, we include Experiment 1 as a demonstration of the inconsistency of cued recognition effects, the likely need for a meaningful link between stimuli to encourage associative encoding, and as a caution while interpreting the findings of Experiment 2. However, we focus primarily on the results of Experiment 2 in order to discuss the results in light of our three research aims. First, our results did not replicate previously-reported cued face recognition effects. Contrary to expectations, in both experiments, cuing did not influence hit rates. However, in Experiment 2, cuing with either correct or incorrect cues (vs. no cues) did reduce false-alarm rates. Therefore, while cuing

enhanced overall accuracy, this occurred through reducing the likelihood of falsely recognising a new face. Second, these effects did extend to scenarios with more faces, such that cuing reduced false alarms in conditions in which participants studied pairs of faces and groups of four faces. Third, our results provide some evidence that cuing helps to reduce the disadvantage of studying multiple faces (vs. individual faces) by reducing response bias. These results are discussed in turn.

Replication of cued face recognition effects

First, we consider whether we could replicate previous effects in cued face recognition. The two-face condition mimicked the original research (Watkins et al., 1976; Winograd & Rivers-Bulkeley, 1977) in cued face recognition, but our methodology differed in two important ways. First, we randomised the left-right placement of the target faces between study and test, such that placement effects could not exert an influence on participant responses. Second, we randomised pairing of faces so that likeness between the faces (i.e., age, impression of personality, etc.) could not provide clues to correct answers. Following cued recognition and the encoding specificity principles (Thomson & Tulving, 1970), we expected in both experiments to see hit rates increase when a correct cue was presented compared with either an incorrect cue or no cue at all. However, this was not the case. Hit rates were not affected by cuing, but false-alarm rates decreased in response to *both* correct and incorrect cuing. In line with these results, we also found response bias decreased when any context was shown, whether it was correct or incorrect context. This means that participants became more conservative in their responses, and were thus less likely to respond "old" when cues were presented. It may be that cuing, regardless of the veracity of the cue, signalled participants to remember to reject in light of uncertainty, or that cuing increased the strength of evidence (i.e., the sense of familiarity or availability of memory)

participants require for a positive response. However, these explanations would need to be explored in future research.

It is interesting to note that the failure to increase hit rates arose in an experiment when face context was isolated from the possibility of placement context and intuitive impressions of couples belonging together. The encoding specificity principle asserts that contextual information is more important than semantic information in cuing (Thomson & Tulving, 1970); If contextual cuing is useful for face recognition, randomised groupings of faces should not theoretically reduce the effect of correct cuing as long as the faces are encoded as context. Therefore, we would expect to still see increased hit rates as a result of correct cuing, and reduced hit rates as a result of incorrect cuing. By contrast, Bayesian analyses for both experiments provided strong evidence that hit rates did not differ as a result of cuing.

It is also interesting to examine these hit rates through the differences in the *pictorial* and *structural* coding present in face recognition (Bruce, 1982). Pictorial coding in face recognition is defined as the description of a static picture of a target face, which is view-specific (i.e., from a single angle or expression), and includes the details specific to the photograph itself (e.g., grain, focus, glare); structural coding is a more abstract representation that retains aspects of the target face that make it distinct from other faces and is therefore robust to changes in deviations (Bruce & Young, 1986). Pictorial coding and structural coding are present, to different degrees, in both actual face recognition and the recognition of pictures of faces (Bruce, 1982). However structural coding becomes more robust the more a face is seen at different views, with different exposures, etc. By using only one photo of a face at study and test, we provide less structural code on which to base recognition decisions. It seems that our face cues, which provided additional pictorial code, were not helpful in retrieving the pictorial code associated with the target face. However, our tests provided less pictorial code than previous tests in cued face recognition (Watkins et al., 1976; Winograd & Rivers-Bulkeley, 1977) because randomised placement of faces did not provide the same physical location of those faces as encoded on the screen during study. Providing more structural code would allow for cues to enhance hit rates.

Instead, the observed cuing-related enhancement in accuracy was a result of reduced false-alarm rates. Thus, cuing did not enhance the true recognition of old faces, but did enhance the ability to reject a new face that was not previously studied. Cued recognition studies typically find that correct cues increase false alarms, but not enough to outweigh the benefits of cuing context (see Davies, 1988). However, our results diverge from previous patterns of contextual cuing, such that false-alarm rates actually decreased as a result of cuing. We should note that it is not unusual for a manipulation in memory research to affect hits and false alarms to different degrees. For example, context reinstatement in eyewitness identification research often inflates the false-alarm rate rather than the hit rate (Shapiro & Penrod, 1986). However, the current research is closely aligned to the previous experiments in cued face recognition. This raises the question why, within the same field of cued recognition research, cues of faces would not present previously-reported risks of increasing false alarms and why it would conversely reduce the false recognition of new faces. Because both correct and incorrect context reduced the false-alarm rate, and because sensitivity was minimally affected by cue type, it would be difficult to argue that memory was enhanced as a result of cuing. However, response bias did become less liberal any time context was present, meaning that participants needed more evidence to report the face as *old*, which was reflected in the reduced false-alarm rate.

Extending cuing effects to more than two faces

Next, we consider whether such cuing effects extend to contexts in which more than two faces are encoded at the same time. This finding was straightforward. While the four-face condition appeared to be at a slight disadvantage compared with the two-face group when no cue was presented, contextual cuing was equally useful in both groups to reduce the false-alarm rate. Signal detection analyses provided further support for this notion, showing that neither sensitivity nor response bias differed between the two- and four-face groups. Thus, cuing effects existed for faces encoded in pairs and groups of four.

Orienting cuing effects

Finally, we attempted to place the cued face recognition effects in context by comparing them to traditional, non-cued recognition memory. This we referred to as orienting the effects of cuing. Given that context reduced false-alarm rates (thus increasing recognition accuracy), we wondered if this is a result of the benefit of providing context (i.e., single faces have no context), or whether it is beneficial only as a means to reduce the disadvantage for having encoded more faces at once. Therefore we compared the no-cue trials across the three face encoding groups. Results are mixed. In Experiment 1, there were no differences in hits or false alarms for test faces, regardless of the number of faces presented at encoding. Contrary to Experiment 1, false-alarm rates in Experiment 2 statistically differed between groups on the no-cue trials such that false-alarm rates in the four-face condition were significantly higher than in the single-face condition, and false-alarm rates in the two-face condition fell between the two. However, Bayesian analyses provided weak support for this difference.

There are a few reasons why we may have obtained these weak differences between groups, such as varying cue types at testing (i.e., the single-face condition was only ever tested with no cues), or the variance of the total number of faces studied at encoding (i.e., the single

face condition studied 36 faces, while the four-face condition studied 36 faces multiplied by 4 blocks). However, we would have expected these issues to arise in Experiment 1, which they did not. Perhaps the most feasible methodological cause for this difference would be that our associative encoding manipulation affected the allocation of attention when participants were encouraged to make judgment tasks. For example, it is possible that participants split their encoding time evenly between pairs of faces, but not groups of four faces. Yet, we would expect such differences in encoding to impact the hit rate, which they did not. What we can say is that something about encoding groups of four faces in our experiment was more difficult for participants than encoding single faces or pairs of faces, even when they encoded the same number of faces in a study-test block. However, this difference was also minimal. Therefore, when participants were given contextual cues, the benefits of cuing appear to be a compensation for what was a slightly more difficult task of seeing multiple faces at once.

Limitations

It is important to consider our results within the limitation that we used identical images at encoding and test. Some researchers justifiably argue that this ignores the natural variability across representations of a face (see Burton, 2013). Because recognition for unfamiliar faces is fragile to even minute deviations, including, lighting, hair-style, image hue, expression, and focal point of the camera (e.g., Jenkins, White, Van Montfort, & Burton, 2011), it is likely that using the same photographs at encoding and test results in an easier task and, thus, overestimates eyewitness memory performance for person recognition (Bruce, 1982). Although it is true that this method does not provide the most realistic test of cued face recognition, there is little reason to assume this approach undermines the results presented here. Research in face recognition using the same images at encoding and test phases has produced similar patterns of results to those tested in the eyewitness paradigm where the faces are always different between the two phases (i.e., confidence-accuracy calibration: Sauer, Brewer, & Weber, 2008; Sauer, Brewer, Zweck, & Weber, 2010; Weber & Brewer, 2004). Interestingly, our results are also in line with Wells and Pozzulo's (2006) test of the novel two-person serial lineup against the traditional simultaneous and sequential lineup procedures: while there were no differences in lineup procedure for accurate identification decisions for target-present lineups, the two-person serial lineup consistently produced fewer false identifications in target-absent lineups. Although our task provides a basis for testing associative memory of faces, future research should vary the photographs used at study and test, or perhaps include a more realistic task (i.e., use of video at study and photos at test). This would allow greater generalizability *to* the more complex real world task of face recognition.

Conclusion

Across two experiments, we sought to replicate previous work that has demonstrated the benefits of cued face recognition for paired faces, to understand those findings in comparison to straightforward single-face recognition, and to extend those findings to situations in which participants study more than two faces. We failed to replicate previous research in cued face recognition with face pairs in the sense that the hit rate for true recognition did not increase when correct cues were available. However, we found that any cue (correct or incorrect) could reduce the false-alarm rate and that these effects extended to those studying groups of four faces, as long as associative encoding was engaged. Furthermore, we demonstrated that cuing likely compensates for the more-difficult task of studying multiple faces at once (i.e., single faces), even when divided attention and memory load are controlled for, and that this occurs by reducing response bias. However, the inconsistent effects reported between Experiments 1 and 2 warrant caution in the application of the findings in real-world contexts (e.g., Wells & Pozzulo, 2006). In order to apply cued face recognition techniques to eyewitness identification procedures, future work should extend such findings in experimental settings that are incrementally closer to the eyewitness identification context, such as using video-to-picture methodology, and using fewer trials (vs. recognition paradigm).

Our results confirm the utility of using other faces as contextual cues to enhance recognition accuracy. However, our work suggests that accuracy is enhanced (1) by a decrease in false recognition rather than an increase in true recognition (decreased false-alarm rate), and (2) a result of a shift in response bias, rather than memorial enhancement. This research is the first replication of original cued face recognition findings using contemporary methodological procedures (i.e., randomisation of face groups and left-right placement), as well as novel research on extending such cued effects to situations in which there are more target faces presented at the same time.

Notes

1. Confidence data for Experiments 1 and 2 are available upon request from first author. Note that due to error in data collection, confidence data were only recorded for choosers.
2. For ease of interpretation, BF_{10} is used to designate the Bayes factor as evidence in favour of the alternative hypothesis and BF_{01} is used to designate the Bayes factor as evidence in favour of the null hypothesis. Note. $BF_{01} = 1/BF_{10}$.

Funding

References

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition*, *98*, 177–197. doi:10.1016/j.cognition.2004.11.004

Bindemann, M., Jenkins, R., & Burton, A. M. (2007). A bottleneck in face identification: Repetition priming from flanker images. *Experimental Psychology*, *54*, 192–201. doi:10.1027/1618-3169.54.3.192

Bindemann, M., Sandford, A., Gillatt, K., Avetisyan, M., & Megreya, A. M. (2012). Recognising faces seen alone or with others: Why are two heads worse than one? *Perception*, *41*, 415–435. doi:10.1068/p6922

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, *103*, 751–757. doi:10.1037/h0037190

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116. doi:10.1111/bjop.1982.73.issue-1

Bruce, V, & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327. doi:10.1111/bjop.1986.77.issue-3

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*, 1467–1485. doi:10.1080/17470218.2013.800125

Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology*, *66*, 364–370. doi:10.1037/0021-9010.66.3.364

Davies, G. D. (1988). Faces and places: Laboratory research on context and face recognition. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 35–53). London: Wiley.

Dempsey, J. L. (2012). *An investigation of retrieval cues as a method to improve eyewitness identification accuracy of multiple perpetrators* (Doctoral dissertation). Carleton University.

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: an eye-tracking study with 3-and 6-month-old infants and adults. *Journal of Experimental Child Psychology*, *113*, 66–77. doi:10.1016/j.jecp.2012.04.012

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology*, *99*, 385–401. doi:10.2307/1422492

Hobson, Z. J., & Wilcock, R. (2011). Eyewitness identification of multiple perpetrators. *International Journal of Police Science & Management*, *13*, 286–296. doi:10.1350/ijps.2011.13.4.253

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*, 2–9. doi:10.7771/1932-6246.1167

JASP Team. (2017). JASP (Version 0.8.1.2)[Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–323. doi:10.1016/j.cognition.2011.08.001

Kan, I. P., Giovanello, K. S., Schnyer, D. M., Makris, N., & Verfaellie, M. (2007). Role of the medial temporal lobes in relational memory: Neuropsychological evidence from a cued recognition paradigm. *Neuropsychologia*, *45*, 2589–2597. doi:10.1016/j.neuropsychologia.2007.03.006

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271. doi:10.1037/0033-295X.87.3.252

Martinez, A. M., & Benavente, R. (1998). *The AR face database* (CVC Technical Report No. 24). Barcelona: Universitat Autonoma de Barcelona, Computer Vision Center.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*, 129–156. doi:10.1037/a0014420

The psychological image collection at Stirling (PICS). (n.d.). University of Stirling Psychology Department. Retrieved from http://pics.psych.stir.ac.uk/.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111. doi:10.2307/271063

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137*, 528–547. doi:10.1037/a0012712

Sauer, J. D., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*, 337–347. doi:10.1007/s10979-009-9192-x

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*, 139–156. doi:10.1037/0033-2909.100.2.139

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220. doi:10.3758/BF03196157

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*, 255–262. doi:10.1037/h0029997

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. doi:10.1037/h0020071

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Meerhoff, F. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*, 58–76. doi:10.3758/s13423-017-1323-7

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior*, *15*, 505–517. doi:10.1016/0022-5371(76)90045-1

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgements. *Journal of Experimental Psychology: Applied*, *10*, 156–172. doi:10.1037/1076-898X.10.3.156

Wells, E. C., & Pozzulo, J. D. (2006). Accuracy of eyewitnesses with a two-culprit crime: Testing a new identification procedure. *Psychology, Crime and Law*, *12*, 417–427. doi:10.1080/10683160500050666

Winograd, E., & Rivers-Bulkeley, N. T. (1977). Effects of changing context on remembering faces. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 397–405. doi:10.1037/0278-7393.3.4.397