The importance of consolidating perceptual experience and contextual knowledge in face recognition

Kira Noad and Timothy J. Andrews

Department of Psychology, University of York, UK.

* Corresponding authors: kn689@york.ac.uk & timothy.andrews@york.ac.uk

Author ORCID identifiers:

Kira Noad: https://orcid.org/0000-0001-8790-2522
Timothy J Andrews: https://orcid.org/0000-0001-8255-9120

Abstract

Although the ability to recognise familiar faces is a critical part of everyday life, the process by which a face becomes familiar in the real world is not fully understood. Previous research has focussed on the importance of perceptual experience. However, in natural viewing, perceptual experience with faces is accompanied by increased knowledge about the person and the context in which they are encountered. Although contextual information is known to be crucial for the formation of new episodic memories, it requires a period of consolidation. It is unclear, however, whether a similar process occurs when we learn new faces. Using a natural viewing paradigm, we will investigate how the context in which events are presented influences our understanding of those events and whether, after a period of consolidation, this has a subsequent effect on face recognition. The context will be manipulated by presenting events in 1) the original sequence, or 2) a scrambled sequence. Although this manipulation is predicted to have a significant effect on the understanding the context of events, it will have no effect on overall visual experience with the faces. Our prediction is that contextual understanding will affect face recognition after the information has been consolidated into memory.

Introduction

Recognising the identity of a familiar face is a straightforward process for most human observers if we are familiar with the person. However, the computational challenge of face recognition becomes apparent when we attempt to recognize people who are less familiar. While familiar face recognition is highly accurate across substantial changes in the image[1–3], unfamiliar face recognition breaks down under small changes in viewing conditions[4–7]. Cognitive models of face perception suggest that we become familiar with a face by generating image-invariant representations[4,5]. During familiarisation, the representation of a face must transition from an image-based representation based on specific encounters into an invariant representation that can be used to recognize the face across different visual environments.

The successful generation of image-invariant representations is thought to depend on perceptual experience whereby different encounters with a face are integrated to create an invariant representation of a facial identity[6,7]. Support for this idea comes from studies that show more visual exposure leads to better recognition of faces[8,9]. A key feature of the familiarisation process appears to be the exposure to the variety of encounters with a person[10,11]. For example, averaged faces made from many different exemplars from the same person are recognised more accurately than faces made from fewer exemplars[12]. These findings provide clear evidence for the importance of visual experience, particularly within-person variability, in becoming familiar with a face.

However, increased perceptual experience is also accompanied by an increase in information about a person (i.e. who they are, what they do, what they are like, where we usually see them) that is distinct from the visual properties of the face. A range of evidence suggests that this information may also play an important role in the generation of invariant representations necessary for familiar face recognition. For example, it has been shown that faces are difficult to recognize in contexts that are different to those in which they are typically encountered[13,14], whereas providing the context in which a face was learnt has been shown to improve face recognition[15–18].

Despite these advances in understanding familiar face recognition, typical paradigms involve viewing a limited number of static images that are associated with arbitrary conceptual knowledge about the person, such as a name or occupation. So, it remains unclear how the recognition of faces unfolds in more naturalistic viewing conditions and over longer time periods. A recent study addressed this issue by measuring face recognition of actors in the TV series Game of Thrones[19]. They found that the faces of the lead actors were recognized better than other actors and that recognition performance was generally better for faces viewed more recently. Although better recognition could reflect increased perceptual experience, it could also reflect increased knowledge about the person.

In natural viewing, we make new memories by integrating information in events or episodes that include what happened, who was present and where and when it happened[20]. A process of consolidation is then necessary if these episodes are to be integrated into longer term memory, which involves the binding, reactivating, and strengthening connections between the hippocampus and distributed neocortical representations[21–23]. Interestingly, this process of acquiring new memories is enhanced when new information is acquired in a coherent context[24,25]. Studies of word learning, for example, show that the successful consolidation of information increases when the words are associated with meaning[26–28]. However, it is not clear whether similar processes are evident in face learning[29–31]. If this is the case, our prediction is that learning faces in a coherent context should lead to more stable recognition over a longer time period compared to when faces are learnt in the absence of a coherent context.

We will use a natural viewing paradigm to explore the effects of perceptual and contextual information on the recognition of faces. Participants who are unfamiliar with the TV series Life on Mars[32], will view excerpts from the series in one of the following conditions: 1) Original sequence or 2) Scrambled sequence. A key feature of the design is that the overall visual input is the same for all the conditions. However, scrambling the sequence will dramatically affect the ability to understand the context or narrative[24,33]. We will then assess whether contextual knowledge has an effect on the recognition of faces. If face recognition is dependent only on visual information, we predict that there should be no difference between any of the conditions. However, if contextual knowledge is important, the recognition of faces will be greater in the Original condition. We will test face recognition immediately after watching the video (short-term) and then 4 weeks later (long-term).

Our preregistered analyses will assess 4 specific hypotheses (each has been assessed in a pilot study). Hypothesis 1: Manipulating the order of the events in the video will affect understanding of the narrative or context. Our prediction is that there will be a greater understanding of the narrative of the stimulus when it is shown in the original sequence compared to a scrambled order. Hypothesis 2: The recognition of faces after a delay will depend on the context in which they were originally presented. Our prediction is that the reduction in face recognition following a delay will be smaller in the Original condition compared to the Scrambled condition, because the greater contextual information in the Original condition will help consolidate the faces in memory. Hypothesis 3: Recognition of faces images will be greater if they are consistent with the appearance at encoding. Our prediction is that face images that are visually similar to the faces at encoding will be recognised to a greater extent compared to images that are not consistent with the appearance at encoding. Hypothesis 4: The effect of context on the recognition of faces, after consolidation, will be greater if the images are consistent with the appearance at encoding. Our prediction is that there will be a bigger difference in recognition scores for In Show compared to Out of Show images for the Original condition compared to the Scrambled condition at the delayed time point.

Methods

Participants

Participants will be recruited who are native English-speaking and are unfamiliar with the TV show Life on Mars[32]. All participants will have either normal or corrected-to-normal vision (by self-report) and will perform the Cambridge Face Memory Test[34] to determine that their face perception is within a normal range (>65%, i.e., not less than 2 SD from the mean). Participants will be compensated with an Amazon voucher or course credit for their time. The study conforms with all relevant ethical regulations at the University of York and was approved by the University of York Department of Psychology Ethics Committee. Informed consent will be obtained from all participants.

Sampling plan

We conducted a sensitivity analysis (see Figure 1) for a one-sided independent t-test with a power of 0.9 and alpha level of 0.02. This showed a rapid initial decrease in the minimum effect size that could be detected, with improvements being relatively marginal beyond around 100 participants per group for our smallest theoretically important effect size (Hypothesis 2: see orange dashed line in Figure 1). We chose this as our sample size, as it allowed us to detect effect sizes of a similar magnitude to that found in our pilot work and also kept the experiment feasible from a practical perspective. This is a 'medium' effect size (see Cohen, 1988), and we consider that effect sizes smaller than this are unlikely to have practical relevance for everyday face recognition performance, so it also constitutes the smallest effect size of interest for this work.

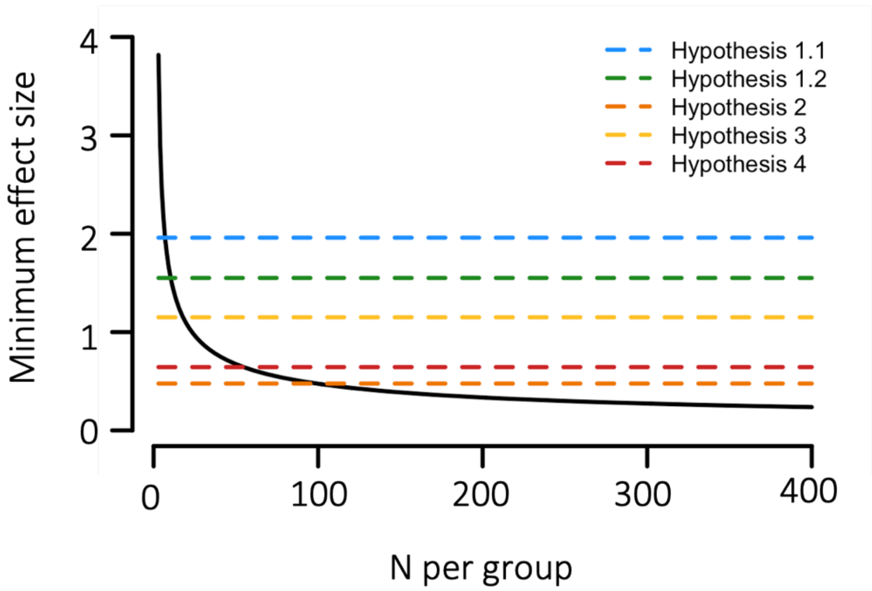


Figure 1. *Sensitivity analysis showing the detectable effect size for a one-sided independent t-test with a power of 0.9 and alpha level of 0.02. The dashed lines represent the effect sizes found in the pilot data for each hypothesis.*

Design

The whole experimental design is 2*2*2 (condition*image type*time). Condition is between-subjects, image type and time are within-subjects.

Stimuli

Two 20-minute (1170s) movies constructed from audio-visual clips from the first episode of BBC TV series Life on Mars[32] will be used as stimuli. Timings are based on previous studies using experimentally familiarised faces[36,37] and on our pilot study. A key aspect of the design is that each movie contains the same visual input. The first video contains the clips in the original order (Original), so that the narrative is coherent. The second video contains the same clips in a randomised order (Scrambled). An illustration of the different video stimuli is shown in Figure 2. A total of 14 clips are used in the stimuli, with a mean length of 84s (range 39s-228s). The clips are assigned a random order for the Scrambled condition, with longer clips cut into shorter segments (mean clip length 39s). 10 unique characters are present in the clips with varying screen time (34-1170s). Participants will be instructed to fully watch and attend to the video before completing any of the other tasks.

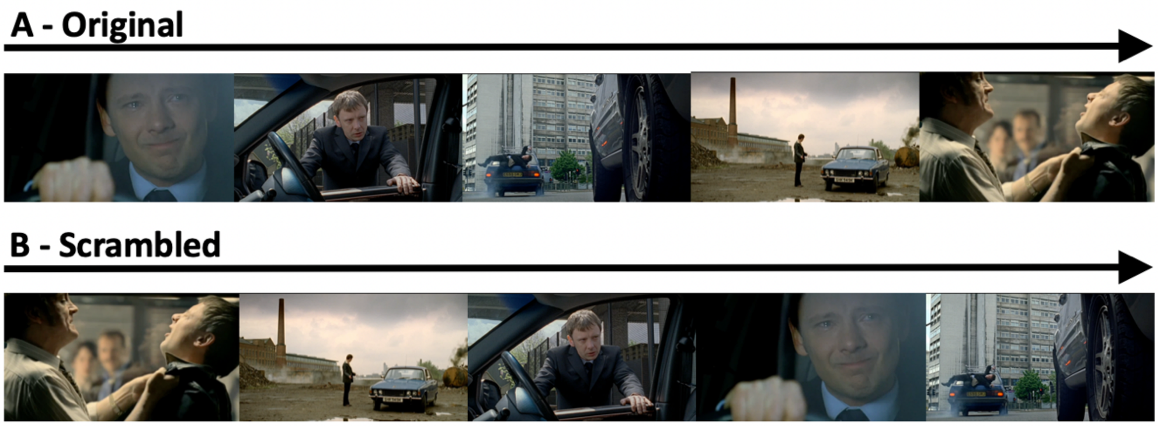


Figure 2. *Illustration of video conditions. The visual exposure is identical, but the order of presentation is different across conditions. (A) The original condition has a narrative that is coherent, while the (B) scrambled condition has an incoherent narrative*

For the face recognition memory task, we will use images from the 10 main actors from the episode of Life on Mars. Static images will be taken directly from the TV series and will be referred to as "In Show" images. However, these will not be images that were seen in the video stimulus. This is critical to avoid confounding face recognition with the visual memory of a specific image[4,5]. Each actor will also have another image from outside of the Life on Mars TV series, which will be referred to as "Out of Show" images. Critically, Out of Show images contain greater within-person variability, with significant changes in physical appearance[6,7]. Previous research has shown that the amount of within-person variability affects subsequent recognition[10,11]. For each In Show or Out of Show face for each face memory test, two foils of different identities will be selected that match the targets in terms of age, expression, hairstyle, lighting, and general appearance[38]. 19 target images (Out of Show image not available for one actor) and 38 foils will be used in each face recognition memory test. Images will be cropped to include the head. Example target and foil images are shown in Figure 3. Different target and foil images will be used at each test phase to avoid practice effects. The comparison between In Show and Out of Show face images will be important to determine whether the effect of context on face memory is specific to the visual context in which the images were originally shown[13,14].

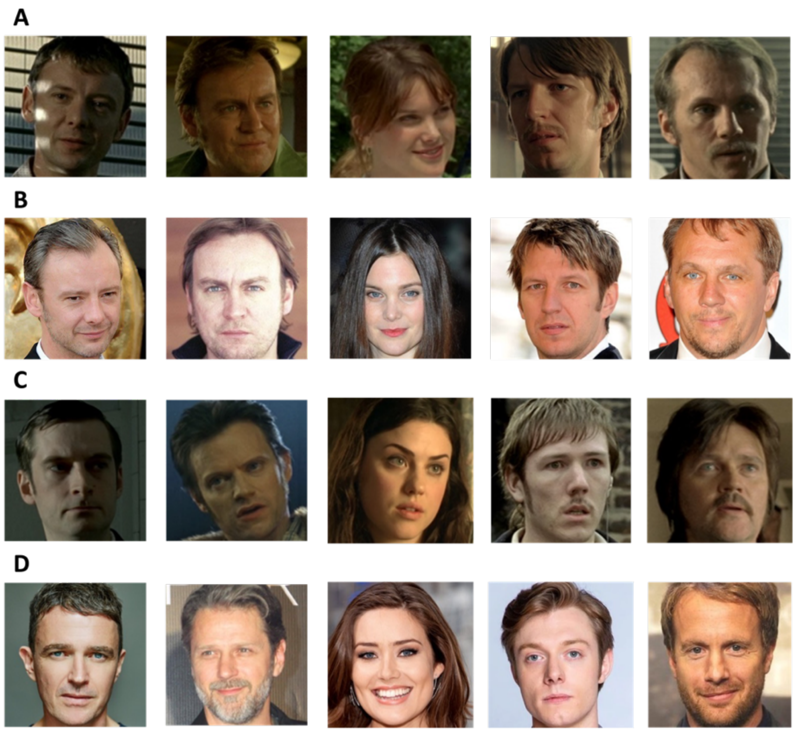


Figure 3. *Examples of faces from the recognition test. The target faces were (A) actors as they appear in the show (In Show) or (B) the actors as they appear out of the show (Out of Show). The foil faces were (C) other actors taken from the same show (In Show foils) or (D) other actors that matched the out of show faces (Out of Show foils).*

Procedure

Participants will be sent a link to a secure website hosting the online experiment. Participants will be prevented from running the experiment on mobile devices. An information sheet will be included with a description of the study, the data that will be collected and how it will be stored, and informed consent will be given. Participants will be randomly allocated to one of the 2 conditions: 1) Original condition, where video clips are viewed in order, or 2) Scrambled condition, where video clips are viewed in a random order.

After being allocated to a condition, participants will commence with the study. During the study phase, participants will be asked to watch and fully attend to the video stimulus. Immediately after the study phase, participants will be tested on their conceptual understanding of the video clips. They will first complete a free recall test, where they will be asked to provide a written description of the plot of the video using as many details as possible. Participants will then complete a face recognition memory task, with faces presented individually in a random order. In this test, participants will press a button to indicate if the identity of the face corresponded to any of the actors in the video. Stimuli will remain on screen until the participant makes a response. Finally, participants will complete a second contextual understanding test (structured question test), containing a series of 8 questions about specific events in the video accompanied by a static image of the relevant event in the video.

A unique participant identifier will be provided by email for participants to complete the face recognition memory task again at 4-weeks after the study phase. A link to the face recognition memory task will be sent at 4-weeks for the participant to access the experiment at the final time-point. For the 4-week time-point the experiment must be completed within 48 hours of the link being sent. Following completion of the study, a debriefing sheet detailing the aims of the experiment will be provided as well as full payment or course credit.

Data analysis

See Table 1 for our study design table with a full list of hypotheses.

Hypothesis 1: Manipulating the order of the events in the video will affect understanding of the narrative or context

Task performance on the conceptual understanding tests will be graded by two raters using a predefined marking scheme. Raters (who are blind to the condition) will mark the free recall test relative to 10 key events that occurred during the video. Raters will assign a mark of 0, 1 or 2 for each point dependent on whether the text showed no, partial or a full description of each event, for a possible total of 20 marks. The 8 questions on the structured question test will also be marked by raters on a scale from 0 to 2, based on whether they show no, partial or a full understanding for a possible total of 16 marks. The analysis will be based on the average scores across raters. Inter-rater reliability will be assessed for both the free recall and structured question test aggregated across questions using intra-class correlation coefficients (ICC) in a two-way mixed model with agreement definition. ICC greater than 0.75 will indicate good reliability between raters. While this value does not need to be achieved for the experiment to be deemed capable of testing the key hypotheses, an ICC greater than 0.75 will validate the marking scheme as effective in consistently assessing the narrative score. The pilot data indicates that reliability should be higher than 0.75.

To assess whether the video manipulation leads to differences in conceptual understanding (Hypothesis 1), the free recall scores (Hypothesis 1.1) and structured question scores (Hypothesis 1.2) for each condition will be entered into a one-tailed independent groups t-test, with an alpha criterion of 0.02 for determining significance. Support for hypothesis 1.1 and 1.2 will be indicated by a significant effect, with lower scores for the Scrambled condition compared to the Original condition. Successful manipulation of video context understanding will be shown if both hypothesis 1.1 and 1.2 are confirmed.

Hypothesis 2: The recognition of faces after a delay will depend on the context in which they were originally presented

Performance on the face recognition memory test will be calculated using the mean sensitivity (*d'*) for discriminating between faces of individuals present in the video and faces of foils who were not present in the video. *d'* will be calculated based on hit rates (i.e. correct recognition of the face as present in the video) and false alarm rates (i.e. incorrectly responding that foil was present in the video) for each participant. *d'* is calculated using the following equation:

$$\boldsymbol{d}' = \mathbf{z}(\mathbf{H}) - \mathbf{z}(\mathbf{FA})$$

where z(H) and z(FA) are the z transforms of the hit rate (number of hits / number of targets) and false alarm (number of false alarms / number of foils), respectively. Ceiling hit rates or false alarm rates (i.e. hit = 1) will be replaced with 0.999 and floor hit rates or false alarm rates (i.e. false alarm = 0) will be replaced with 0.001 to avoid *d'* infinity. *d'* will be calculated separately for each face recognition memory test time point (0 hours, 4-weeks) and separately for In Show face images and Out of Show face images.

To determine if contextual understanding has a role in recognition of faces after a delay in stimulus encoding (Hypothesis 2), the difference between the immediate and delayed (immediate – delayed) face recognition score (*d'*) for each condition (Original and Scrambled) will be calculated separately and then compared using a one-tailed independent groups t-test for the In Show images. Support for hypothesis 2 will be shown if the difference scores are greater in the Scrambled compared to the Original condition at $p<.02$.

Hypothesis 3: Recognition of face images will be greater if they are consistent with the appearance at encoding

The average score (*d'*) will be combined across timepoints for the In Show and Out of Show images in the Original condition. To determine whether the appearance of the images at encoding is important for subsequent recognition, a one-tailed independent groups t-test will be performed on the difference between In Show and Out of Show face recognition. Support for hypothesis 3 will be indicated by a greater face recognition score for In Show images than Out of Show images at $p<.02$.

Hypothesis 4: The effect of context on the recognition of faces, after consolidation, will be greater if the images are consistent with the appearance at encoding

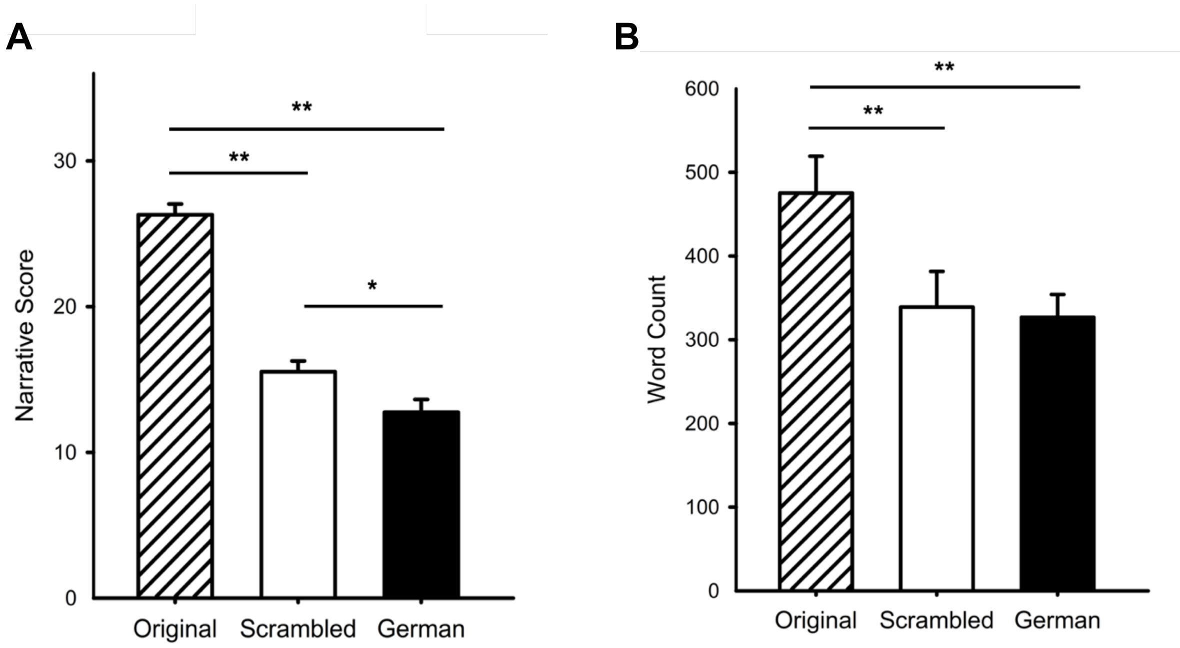
To investigate whether the role of contextual understanding on face recognition after consolidation is influenced by the appearance of the faces at encoding, the difference between the recognition for In Show and Out of Show images (In Show – Out of Show) will be calculated for each condition (Original, Scrambled) and compared using a one-tailed independent groups t-test at the delayed time point. Support for hypothesis 4 will be indicated by a bigger difference in face recognition between In Show images compared to Out of Show images for the Original condition at $p<.02$.

Exclusion Criteria

Participants who do not complete the face recognition test at all time points will be excluded from all analyses; participants who do complete the delayed face recognition test but not within the specified time slot will also be excluded from analysis. Participants who do not complete both the free recall and structured narrative questions will be excluded from analysis. Participants will be asked at each time point if they have seen the TV show Life on Mars; participants who have seen the show at any point will be excluded from all analyses. Participants will be screened for familiarity with other popular shows characters have been in, such as Ashes to Ashes (2008) which shares characters and actors with Life on Mars. They will also be excluded if they have seen the TV show Spooks (2002), as foil images were gathered from this show.

Pilot Study

We performed an exploratory online pilot study with 156 (82 completed both time points) participants to investigate the interaction between perceptual and conceptual information in face recognition. The results validate the experimental design for the main experiment and generate estimated effect sizes for a sample size computation. The methods and design are identical to the proposed study, except an additional condition was included. Participants in the additional German condition watched the video clips in the original order but with the soundtrack dubbed in German, meaning that the narrative was incoherent.

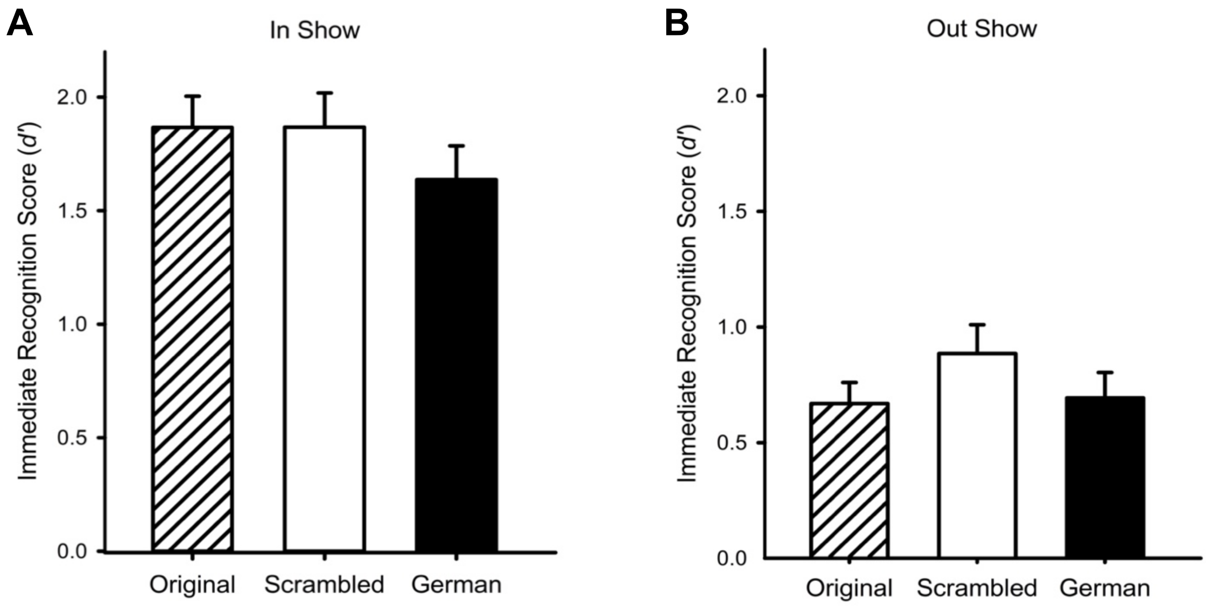


Suppl. Fig. 1. *Performance on narrative tests of contextual understanding across the different experimental conditions in pilot study (A) and difference in word count in the free recall task between conditions (B). This shows that contextual understanding was greater in the Original condition compard to the Scrambled or German conditions. Error bars indicate standard error. ** p < .001, * p < .05*

Suppl. Fig. 1 shows the differences in contextual understanding across the different conditions (Original, Scrambled, German). A one-way ANOVA revealed a significant difference in narrative score between the conditions ($F(2,153) = 88.82$, $p <.001$, $\eta^2 = 0.54$). Planned comparisons showed that the narrative score for the Original condition was significantly greater than for the Scrambled ($p < .001$), and the German condition ($p < .001$). The Scrambled and German group were significantly different ($p = .013$). Agreement between raters for the conceptual understanding tests was calculated using intra-class correlation coefficient (ICC) with a two-way mixed model and Agreement definition. Excellent agreement was found between raters in the free recall test with an ICC of .91 and 95% confidence intervals of .88 to .94 ($F(155,155) = 22$, $p < .001$) and in the structured question test with an ICC of .93 and 95% confidence intervals of .90 to .95 ($F(155,155) = 27$, $p < .001$). We also measured the total word count across the conditions in the Free Recall task. A one-way ANOVA revealed significant differences in word count between the conditions ($F(2,153) = 4.77$, $p = .010$, $\eta^2 = 0.06$). The mean score for the Original condition was significantly higher than for the Scrambled ($p = .012$) or German ($p =$

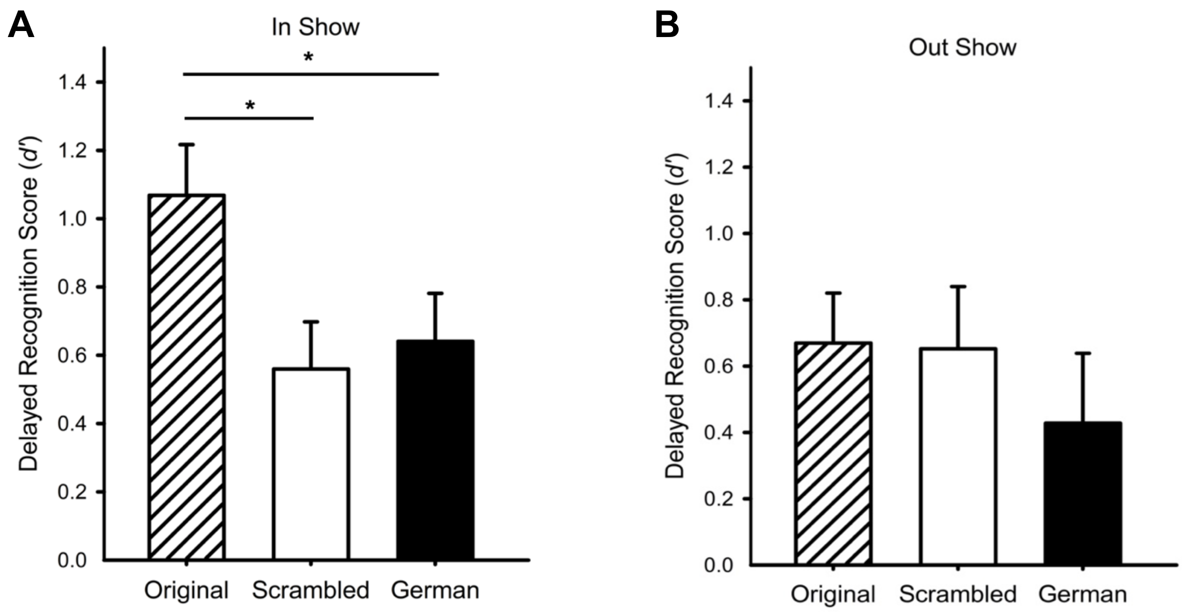
.006) conditions. This indicates a more detailed response in the Original condition. There was no significant difference between the Scrambled and German group in mean word count (*p* = .827). Together, these findings show that the contextual understanding of the stimulus was greatest in the Original condition.

Next, we investigated the extent to which differences in contextual understanding affected face recognition. First, we measured face recognition immediately after participants had watched the video. Suppl. Fig. 2 shows differences in recognition across the different conditions. A one-way ANOVA showed that there was no significant effect of condition on the In Show ($F(2,153) = .901$, $p = .409$, $\eta^2 = 0.01$) or Out of Show ($F(2,153) = 1.25$, $p = .288$, $\eta^2 = 0.02$) faces. This implies that differences in contextual information did not influence the recognition of faces. Rather, it would appear that recognition is based on the overall visual or perceptual experience of the faces.



Suppl. Fig. 2. *Immediate face recognition performance (d') for each condition for pilot study. (A) shows d' for In Show faces stimuli and (B) shows d' for Out of Show faces. Error bars indicate standard error.*

We then repeated the face memory task after approximately 4-weeks to explore how the consolidation of the events influenced face recognition. Although all of the 156 participants were invited to take part, only 82 completed the task (Original N = 27, Scrambled N = 27, German N = 28). The average delay between the immediate and delayed tasks was 65.4 $\pm$ 33.2 days, which was equivalent across groups. In contrast to the pre-registered study, the same images were used in both the immediate and delayed face recognition tests. Suppl. Fig. 3 shows the mean face recognition scores for each condition. There was a significant difference in the recognition of In Show faces across the conditions (F(2,79) = 4.51, p = .014, η^2 = 0.103). This difference for In Show faces reflects a significant difference between the Original condition and both the Scrambled (p = .005) and German (p = .032) conditions. There was no difference between the Scrambled and German conditions for In Show faces (p = .474). In contrast, there was no effect of condition for the Out of Show faces (F(2,79) = .58, p = .560, η^2 = 0.016). This suggests a consolidation of contextual and perceptual knowledge leads to increased face recognition. However, this is only evident for faces from the In Show condition. This suggests that the effect of consolidation is restricted to face images that are similar to those seen at encoding.



Suppl. Fig. 3. *Long-term face recognition performance (d') for each condition for pilot study. (A) shows d' for In Show faces stimuli and (B) shows d' for Out of Show faces. Error bars are standard errors. * p < .05*

To investigate the role of contextual information and consolidation on subsequent face recognition, two-tailed independent groups t-tests were run on the difference in face recognition score at the immediate and delayed time-point for the In Show and Out of Show images separately. There was no significant difference in recognition of In Show faces after a delay between the Original (M = 0.81, SD = 1.03) and German (M = 0.95, SD = 1.01) groups (t(52.8) = 0.52, p = .608, d = 0.139). There was no significant difference in recognition of Out of Show faces after a delay between the Original (M = -0.02, SD = 0.79) and Scrambled (M = 0.27, SD = 0.97) groups (t(49.9) = 1.2, p = .235, d = 0.327) or the Original and German (M = 0.07, SD = 1.06) groups (t(49.9) = 0.36, p = .723, d = 0.095). However, there was a near significant difference in recognition of In Show faces after a delay between the Original and Scrambled (M = 1.3, SD = 1.03) groups (t(52) = 1.75, p = .086, d = 0.476), suggesting that understanding of contextual context and consolidation under some conditions may be important for subsequent recognition of faces with a similar visual context.

To investigate how the appearance of the image at encoding affects face recognition, one-tailed independent groups t-tests were run on the In Show and Out of Show average d prime scores over both time points for the Original condition and Scrambled condition separately. A significant difference between the In Show (M = 1.47, SD = 0.75) and Out of Show (M = 0.66, SD = 0.67) images was found for the Original condition (t(51.3) = 4.23, p < .001, d = 1.150). Similarly, a significant difference between the In Show (M = 1.15, SD = 0.67) and Out of Show (M = 0.79, SD = 0.69) images was found for the Scrambled condition (t(51.9) = 1.9755, p = .028, d = 0.538). This is consistent with the idea that suggests face memory is dependent on the specific perceptual experience[10,11].

To investigate if, after consolidation, contextual information has a greater effect on recognition of faces that are similar to the appearance at encoding, two-tailed independent groups t-tests were run on the difference in face recognition scores for the In Show and Out of Show images for each condition at the immediate and delayed time-point. There was no significant difference in recognition of In Show and Out of Show faces at the immediate time point between the Original (M = 1.23, SD = 1.22) and German (M = 1.1, SD = 0.97) conditions (t(49.6) = 0.45, p = .656, d = 0.122) or between the Original and Scrambled (M = 0.88, SD = 1.1) conditions (t(51.4) = 1.10, p = .276, d = 0.299). There was no significant difference in recognition of In Show and Out of Show faces at the delayed time point between the Original (M = 0.4, SD = 0.92) and German (M = 0.21, SD = 1.21) conditions (t(50.3) = 0.64, p = .523, d = 0.173). However, there was a significant difference in recognition of In Show and Out of Show faces at the delayed time point between the Original and Scrambled (M = -0.15, SD = 0.78) group at the delayed time point (t(50.5) = 2.37, p = .022, d = 0.644), showing that recognition of In Show faces was greater than Out of Show faces after a delay for the Original group compared to the Scrambled group. This shows the importance of consolidation of contextual information for recognition of faces, but only for faces that are similar in appearance to those at encoding.

Data availability

All anonymised behavioural data collected via the online task will be made freely available on the Open Science Framework (OSF) website.

Code availability

All analysis code will be made freely available on the OSF website.

Competing interests

The authors declare no competing interests.

Table 1. Hypotheses, sampling and analysis plan for each of the research questions

Hypothesis		Sampling plan	Analysis Plan	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
1. Manipulating the order of the events in the video will affect understanding of the narrative or context	1.1. A greater understanding of the narrative as shown in the free recall test will be found in the Original condition than the Scrambled condition.	A sensitivity analysis has been computed for a one-sided independent t-test with a power of 0.9 and alpha level of 0.02. Our proposed sample size of 100 participants per condition (Original and Scrambled) is sufficient to detect the effect size for hypothesis 1.1 (d = 1.96 [upper CI: 2.43, lower CI: 1.49]), for hypothesis 1.2 (d = 1.55 [upper CI: 1.99, lower CI: 1.12]) and for hypothesis 2 (d = 0.476 [upper CI: 1.03, lower CI: -0.078]) found in the pilot study (see figure 1).	A one-tailed independent groups t-test will be performed to determine if conceptual understanding scores in the free recall test are greater in the Original compared to the Scrambled condition. A statistical effect of condition at $p<.02$ will confirm hypothesis 1.1.	A significant effect of condition would show that the video manipulation has successfully affected context understanding. A lower score for the Scrambled condition compared to the Original condition will show that a coherent context leads to better understanding. Successful manipulation of video context understanding will be concluded only if both hypothesis 1.1 and 1.2 are confirmed. A higher score for the Scrambled condition compared to the Original condition would disconfirm the hypothesis.	A null result would suggest that the video manipulation did not affect context understanding as intended. The outcome of experimental hypotheses 2 and 4 cannot be interpreted as due to the effects of context.
	1.2. A greater understanding of the narrative as shown in the structured question test will be found in the Original condition than the Scrambled condition.		A one-tailed independent groups t-test will be performed to determine if the structured question test scores are greater in for the Original compared to the Scrambled conditions. A statistical effect of condition at $p<.02$ will confirm hypothesis 1.2.		
2. The recognition of faces after a delay will depend on the context in which they were originally presented			The difference between the immediate and delayed face recognition scores (immediate – delayed) will be calculated separately for the Original and Scrambled conditions using the In Show images. A one-tailed independent groups t-test will then be performed to determine if the difference scores are greater in the Scrambled compared to the Original condition. A statistical	A larger difference in face recognition between the immediate and delayed timepoints in the Scrambled compared to the Original condition would show that the context in which a face is presented influences subsequent recognition, if it has been consolidated into memory. If there is no difference or a larger difference in recognition in the Original compared to the Scrambled condition, then this would disconfirm the hypothesis.	A null result for In Show images would suggest that consolidation of contextual information does not have a role in face learning and recognition. This would challenge models of memory consolidation[25], which posit that consolidation is necessary.

		difference at $p<.02$ will confirm hypothesis 2.		for extraction of contextual information into memory. A significant result would support the role of contextual understanding and consolidation in face learning, which is not central to current models of face learning.
3. Recognition of faces images will be greater if they are consistent with the appearance at encoding	Based on the sensitivity analysis (see figure 1), our proposed sample size of 100 participants per condition (Original and Scrambled) is sufficient to detect the effect size found in the pilot for hypothesis 3 (d = 1.15 [upper CI: 1.7, lower CI: 0.56]).	Recognition of In Show faces and Out of Show faces will be averaged across the immediate and delayed timepoints in the Original condition. A one-tailed independent groups t-test will be performed to assess if recognition of In Show faces is greater than Out of Show faces. A statistical effect of condition at $p<.02$ will confirm hypothesis 3.	Higher recognition of In Show compared to Out of Show images would show that face memory is dependent on visual experience at encoding. A non-significant effect or greater recognition of Out of Show images than In Show images would disconfirm hypothesis 3.	A significant result would support theory that suggests face memory is dependent on perceptual experience at encoding [10, 11]. A null result would challenge current theory about the importance of experience with within-person variability in face recognition.
4. The effect of context on the recognition of faces after consolidation will be greater if the images are consistent with the appearance at encoding	Based on the sensitivity analysis (see figure 1), our proposed sample size of 100 participants per condition (Original and Scrambled) is	The difference between In Show and Out of Show face recognition scores (In Show – Out of Show) will be calculated for the Original and Scrambled groups at the delayed timepoint. A one-tailed independent groups t-test will determine if the difference scores in the Original condition are greater than in the	A greater difference between In Show and Out of Show face recognition in the Original compared to the Scrambled condition at the delayed timepoint would show that the effect of context on face recognition following consolidation is greater when the face images are consistent with the way that they appeared at encoding. A non-significant result would imply that the effect of context on face recognition following	A null result would challenge current theory about the importance of experience with within-person variability in face recognition. A significant result would support theory that suggests face memory is

	sufficient to detect the effect size found in the pilot for hypothesis 4 (d = 0.64 [upper CI: 1.2, lower CI: 0.08]).	Scrambled condition. A statistical effect at *p*<.02 will confirm hypothesis 4.	consolidation does not depend on whether the face images are consistent with the appearance at encoding.	dependent on perceptual experience at encoding [10,11], and support the role of contextual understanding and consolidation in face learning, which is not central to current models of face learning.

References

1. Bruce, V. Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology* **73**, (1982).
2. Burton, A. M., Wilson, S., Cowan, M. & Bruce, V. Face recognition in poor-quality video: Evidence from Security Surveillance. *Psychol Sci* **10**, (1999).
3. Mike Burton, A. Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology* **66**, (2013).
4. Bruce, V. & Young, A. Understanding face recognition. *British Journal of Psychology* **77**, (1986).
5. Young, A. W. & Burton, A. M. Recognizing Faces. *Curr Dir Psychol Sci* **26**, (2017).
6. Burton, A. M., Jenkins, R. & Schweinberger, S. R. Mental representations of familiar faces. *British Journal of Psychology* **102**, (2011).
7. Kramer, R. S. S., Young, A. W. & Burton, A. M. Understanding face familiarity. *Cognition* **172**, (2018).
8. Memon, A., Hope, L. & Bull, R. Exposure durations: Effects on eyewitness accuracy and confidence. *British Journal of Psychology* vol. 94 Preprint at https://doi.org/10.1348/000712603767876262 (2003).
9. Roark, D. A., Otoole, A. J., Abdi, H. & Barrett, S. E. Learning the moves: The effect of familiarity and facial motion on person recognition across large changes in viewing format. *Perception* **35**, (2006).
10. Juncu, S., Blank, H., Fitzgerald, R. J. & Hope, L. Do Image Variability and Names in Missing Person Appeals Improve Prospective Person Memory? *J Appl Res Mem Cogn* **9**, (2020).
11. Ritchie, K. L. & Burton, A. M. Learning faces from variability. *Quarterly Journal of Experimental Psychology* **70**, (2017).
12. Burton, A. M., Jenkins, R., Hancock, P. J. B. & White, D. Robust representations for face recognition: The power of averages. *Cogn Psychol* **51**, (2005).
13. Thomson, D. M. Face Recognition: More Than a Feeling of Familiarity? in *Aspects of Face Processing* (1986). doi:10.1007/978-94-009-4420-6_11.
14. Young, A. W., Hay, D. C. & Ellis, A. W. The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology* **76**, (1985).
15. Hanczakowski, M., Zawadzka, K. & Macken, B. Continued effects of context reinstatement in recognition. *Mem Cognit* **43**, (2015).

16. Reder, L. M. *et al.* Why It's Easier to Remember Seeing a Face We Already Know Than One We Don't: Preexisting Memory Representations Facilitate Memory Formation. *Psychol Sci* **24**, (2013).
17. McCrackin, S. D., Lee, C. M., Itier, R. J. & Fernandes, M. A. Meaningful faces: Self-relevance of semantic context in an initial social encounter improves later face recognition. *Psychon Bull Rev* **28**, (2021).
18. Schwartz, L. & Yovel, G. The roles of perceptual and conceptual information in face recognition. *J Exp Psychol Gen* **145**, (2016).
19. Devue, C., Wride, A. & Grimshaw, G. M. New insights on real-world human face recognition. *J Exp Psychol Gen* **148**, (2019).
20. Tulving, E. Episodic memory: From mind to brain. *Annu Rev Psychol* **53**, (2002).
21. Squire, L. R. & Zola-Morgan, S. The medial temporal lobe memory system. *Science* vol. 253 Preprint at https://doi.org/10.1126/science.1896849 (1991).
22. Nadel, L. & Moscovitch, M. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* **7**, (1997).
23. Yonelinas, A. P., Ranganath, C., Ekstrom, A. D. & Wiltgen, B. J. A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews Neuroscience* vol. 20 Preprint at https://doi.org/10.1038/s41583-019-0150-4 (2019).
24. van Kesteren, M. T. R., Fernández, G., Norris, D. G. & Hermans, E. J. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc Natl Acad Sci U S A* **107**, (2010).
25. Lewis, P. A. & Durrant, S. J. Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences* vol. 15 Preprint at https://doi.org/10.1016/j.tics.2011.06.004 (2011).
26. Davis, M. H. & Gaskell, M. G. A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 364 Preprint at https://doi.org/10.1098/rstb.2009.0111 (2009).
27. Henderson, L., Devine, K., Weighall, A. & Gaskell, G. When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Dev Psychol* **51**, (2015).
28. Williams, S. E. & Horst, J. S. Goodnight book: Sleep consolidation improves word learning via storybooks. *Front Psychol* **5**, (2014).
29. Bird, C. M. & Burgess, N. The Hippocampus Supports Recognition Memory for Familiar Words but Not Unfamiliar Faces. *Current Biology* **18**, (2008).

30. Olsen, R. K. *et al.* The role of relational binding in item memory: Evidence from face recognition in a case of developmental amnesia. *Journal of Neuroscience* **35**, (2015).

31. Mattarozzi, K., Colonnello, V., Russo, P. M. & Todorov, A. Person information facilitates memory for face identity. *Psychol Res* **83**, (2019).

32. Life on Mars. (2006).

33. Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S. & Reynolds, J. R. Event perception: A mind-brain perspective. *Psychological Bulletin* vol. 133 Preprint at https://doi.org/10.1037/0033-2909.133.2.273 (2007).

34. Duchaine, B. & Nakayama, K. The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia* **44**, (2006).

35. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* **39**, 175–191 (2007).

36. Hahn, C. A. & O'Toole, A. J. Recognizing approaching walkers: Neural decoding of person familiarity in cortical areas responsive to faces, bodies, and biological motion. *Neuroimage* **146**, (2017).

37. Hahn, C. A., O'Toole, A. J. & Phillips, P. J. Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology* **107**, (2016).

38. Colloff, M. F., Wilson, B. M., Seale-Carlisle, T. M. & Wixted, J. T. Optimizing the selection of fillers in police lineups. *Proc Natl Acad Sci U S A* **118**, (2021).