This version of the manuscript has not gone under review

Context reexposure to bolster contextual dependency of emotional episodic memory

Wouter R. Cox[1*], Mandy Woelk[2], Olivier T. de Vries[1], Angelos-Miltiadis Krypotos[3], Merel Kindt[1], Iris M. Engelhard[3], Dieuwke Sevenster[3], Vanessa A. van Ast[1*]

[1]Department of Clinical Psychology, University of Amsterdam, The Netherlands
[2]Research Unit Behaviour, Health, and Psychopathology, KU Leuven, Belgium
[3]Department of Clinical Psychology, Utrecht University, the Netherlands

Correspondence to: W.R.Cox@uva.nl, V.A.vanAst@uva.nl

ABSTRACT

Contextual overgeneralization of emotional memory is believed to be a core aspect of affective disorders. Identifying methods to restrict emotional memory activation to its original encoding context is therefore of significant clinical interest. Preliminary evidence from rodent research points to a promising approach: reexposure to the context in which fear conditioning took place seems to reduce fear generalization to other contexts. However, it remains unknown which mechanisms underlie these effects, even though such fundamental knowledge is crucial for successful translation to interventions. Importantly, exposure to a context that resembles – but is not identical to – the encoding context may lead to diminished contextual dependency of memory by integration of additional contextual cues. Here, we therefore assessed in a large-scale study (*N*=180) whether reexposure to the encoding context enhances contextual dependency of episodic memory and whether exposure to a similar context impairs it. We also tested whether such effects are predicted by the strength of memory retrieval during context (re)exposure. The results showed that correct recognition depends on context and that such contextual dependency is lower for emotional than neutral memories, which replicates prior research. However, exposure to the encoding context or a similar context did not affect contextual dependency of emotional or neutral memory. Also, retrieval strength did not predict any effects. Thorough insight into factors underlying the effects of context (re)exposure on contextual dependency seems key to eventually attain a therapeutic memory recontextualization intervention.

1. INTRODUCTION

When one returns to the spatial environment in which an event has previously occurred, retrieval and recognition of that event often happens readily, whereas memories can be more difficult to reach in different contexts (Godden & Baddeley, 1975; Smith, 1979; Smith & Vela, 2001). Such contextual dependency of memory is, however, less pronounced for emotional compared to neutral memories, an observation that is partly attributed to relatively weak embedding of emotional events in their encoding context (Cox, Meeter, et al., 2022; van Ast et al., 2013, 2014). The reduced contextual dependency of emotional memory is believed to lie at the heart of dysfunctional memory development, resulting in symptoms that characterize a multitude of affective disorders (Maren et al., 2013). Influential clinical models, for example, posit that little processing of the context in which a trauma took place makes resulting emotional memories prone to reactivation by trauma-related cues in safe contexts. This may manifest itself in symptoms that are characteristic of anxiety disorders and posttraumatic stress disorder (PTSD) such as fear generalization and intrusive images (Acheson et al., 2012; Brewin et al., 2010; Ehlers & Clark, 2000; Lambert & McLaughlin, 2019; Liberzon & Abelson, 2016). In support of this idea, experimental studies have demonstrated that reductions in contextual dependency of memory indeed predict the occurrence or distress of analogue trauma intrusions (Bisby et al., 2010; Meyer et al., 2017; Voorendonk et al., 2021). Insights into how contextual dependency of emotional memory can be targeted is therefore of significant clinical interest.

Neuroscientific research points to the hippocampus – a brain region well-known for its crucial role in episodic memory (Eichenbaum, 2004) – as serving several functions that can either foster or distort contextual dependency of emotional events (Desmedt et al., 2015; Maren et al., 2013). A subregion of the hippocampus, the dentate gyrus, is involved in orthogonalizing a memory relative to similar memories, such that it becomes more distinct, ensuring that the memory remains exclusively linked to its context (i.e., pattern separation (Yassa & Stark, 2011). Other subregions, such as area CA1, are believed to aid in binding individual elements of a single experience (Dimsdale-Zucker et al., 2018). Hence, proper integration of events in their encoding context involves increased discrimination in the dentate gyrus and stronger binding in the hippocampal CA1 area, together likely protecting against memory retrieval by resembling experiences. How can involvement of these regions be promoted such that contextual dependency of memory is *retrospectively* enhanced? Several fear-conditioning studies have identified a promising method to this end: reexposure to the spatial context in which encoding took place seems to neutralize later fear responses in other contexts (Al Abed et al., 2020; de Oliveira Alvares et al., 2012; Sekeres et al., 2020; Sevenster et al., 2017, 2018; Wiltgen & Silva, 2007; Winocur et al., 2009; Zhou & Riccio, 1994). Notably, brain regions involved in integration, such as hippocampal area CA1 (Al Abed et al., 2020), have been found to mediate such effects. It thus seems that integration of contextual information in existing emotional memory may facilitate the enhancement of contextual dependency by context reexposure.

Interestingly, interventions that target anxiety disorders and PTSD may partly work through this process. Indeed, contextual processing has been proposed as an important element of effective treatment (Ehlers & Clark, 2009) and interventions such as imaginal exposure (Foa et al., 1999) and imagery rescripting (Holmes et al., 2007) involve the (imagined) revisiting of the trauma context (in addition to other crucial elements, e.g. Kunze et al., 2019). A thorough understanding of these processes in relation to context reexposure may be imperative to ultimately improve treatment in clinical practice. For example, when executing in vivo exposure for anxiety disorders, a return to the same context as during the emotional event is very often not possible (e.g., in cases of impassable or untraceable environments related to war traumas). Clinicians may, therefore, be inclined to expose their patients to contexts that perceptually resemble the encoding context as an alternative to reach comparable therapeutic effects. Similarly, when applying imagery techniques, a shallow imagination of the traumatic event may not lead to reinstatement of the encoding context but result in reactivation of a different or degraded context representation. Provided that memory integration is at play, it is doubtful whether desirable treatment outcomes would then be reached. This is because binding of the original emotional memory with a context that resembles the encoding context could be triggered, which may result in a more *generic* – instead of specific - memory representation. This could in turn generate *impaired* – not enhanced – contextual dependency of memory (Sevenster et al., 2017). In line with this idea, rodent research suggests that when an animal is exposed to a context that is similar to the conditioning context, fear generalization is subsequently amplified (de Oliveira Alvares et al., 2012; Fujinaka et al., 2016). Thus, exposure to a similar context could inadvertently be counterproductive. Knowing if and when such effects occur is critical to promote favorable treatment outcomes.

Mechanistic insight is likewise needed to know the optimal conditions for successful recontextualization of emotional memory when (imagined) revisiting of the original encoding context is in fact achieved. If integration of contextual information with the existing emotional memory underlies this therapeutic effect, strong memory retrieval during context reexposure is probably required. This can be expected as previous research suggests that memory reactivation of an earlier event drives the integration of additional information (Cox, Dobbelaar, et al., 2021; van Kesteren et al., 2018, 2020; Zeithamova et al., 2012). Based on these insights, it thus seems likely that stronger retrieval during context reexposure predicts larger increases in contextual dependency. This may especially be true for emotional memories that typically depend less on context than neutral memories (Cox, Meeter, et al., 2022; van Ast et al., 2013, 2014). Stronger retrieval during exposure to a *similar* context may also amplify memory integration, but in this case of the original learning event with a different context (Josselyn & Frankland, 2018). Because integration of contextually dissimilar memories may reduce the specificity of memory (Sevenster et al., 2017), stronger memory retrieval should then predict smaller contextual dependency

In the present study, we thus aimed to test whether (i) reexposure to the encoding context improves contextual dependency of episodic memory and (ii) exposure to a similar context impairs it. Furthermore, we tested if both effects (iii) become more pronounced when memory retrieval is stronger during context (re)exposure and (iv) are different in magnitude for emotional versus neutral memories. Participants (*N*=180) were shown a series of images of faces (neutral faces: *N*=90, angry faces: *N*=90) on background pictures (spatial contexts). One day later, the two groups were divided in three subgroups (*N*=30) that were either (i) reexposed to the spatial contexts seen on day 1 (Same Context Exposure group), (ii) were shown similar contexts (Similar Context Exposure group), or (iii) did not come to the laboratory (No Exposure group). To stimulate memory retrieval and measure its strength, participants that were exposed to contexts (Same Context Exposure group or Similar Context Exposure group) were instructed to indicate to what extent they relived the episodes they had encoded on the first day. On the third and final day, all participants underwent a recognition test during which they were presented with targets (i.e., faces seen during encoding) and lures (i.e., unfamiliar faces). As a test of contextual dependency of memory, one third of the targets and lures were presented on one of the contexts that were seen on the first day, i.e., the same context for targets and random contexts for lures (Test in Original Context condition). Another third of the faces was shown on a similar context (Test in Similar Context condition) and the final third was shown on a completely new context (Test in New Context condition). Generally, context-dependent memory should be reflected by most accurate performance when testing takes place in the original encoding context, followed by testing in a similar context, and finally a new context (see Fig. 1, right pane for a graphical depiction of this rationale). To assess changes in this contextual dependency by context (re)exposure, we calculated two difference scores: Test in Original Context – Test in New Context (Contextualization Original Context score) and Test in Similar Context – Test in New Context (Contextualization Similar Context score). An *increase* in contextual dependency should be expressed as an enhanced Contextualization Original Context score, together with a decreased Contextualization Similar Context score (i.e., if contextual dependency is bolstered, memory should be highly specific for the original encoding context only, Fig. 1,

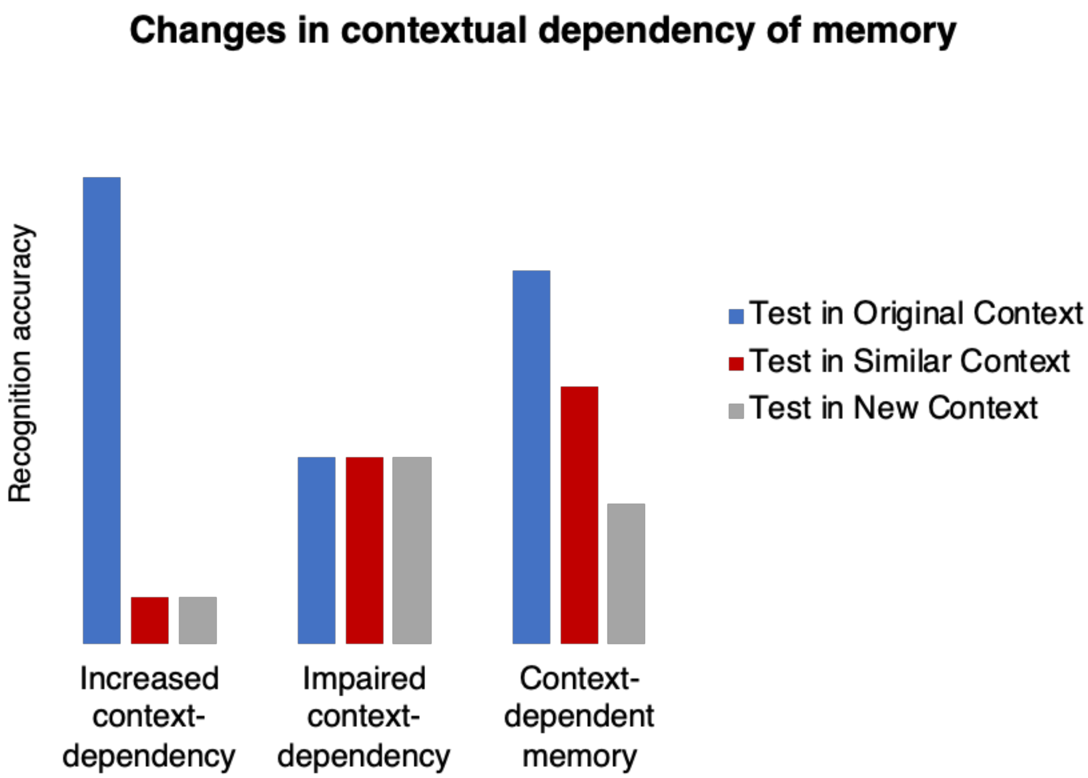

Fig. 1. Illustration of increased and impaired contextual dependency of episodic memory (fictitious data). Generally, context-dependent memory (right pane) would be expressed as most accurate memory when testing takes place in the original encoding context (Test in Original Context), followed by a similar context (Test in Similar Context), and finally a new context (Test in New Context). An increase in this overall contextual dependency of memory (left) should be expressed as memory becoming highly specific for the original encoding context only. That is, the difference in performance between Test in Original Context and Test in New Context (Contextualization Same Context) would be enhanced, whereas the difference between Test in Similar Context and Test in New Context (Contextualization Similar Context) would be decreased. Impaired contextual dependency of memory (middle pane) should be expressed as memory becoming more equal for testing in original, similar, and new contexts. Hence, the difference in performance between Test in Original Context and Test in New Context (Contextualization Same Context), as well as Test in Similar Context and Test in New Context (Contextualization Similar Context) should become relatively decreased.

left pane). *Impaired* contextual dependency should be reflected as a decreased Contextualization Original Context score, together with a decreased Contextualization Similar Context score (i.e., if contextual dependency is impaired, memory should become roughly equal in original, similar, and new contexts, Fig. 1, middle pane). We thus predicted that relative to the No Exposure group, participants in the Same Context Exposure group would show both a higher Contextualization Original Context score and a smaller Contextualization Similar Context score, whereas participants in the Similar Context Exposure group would show smaller scores for both Contextualization Original Context and Contextualization Similar Context. Furthermore, we expected that these effects would become more pronounced when memories were relived to a larger degree during context (re)exposure on day 2. Finally, as emotional memories should overall be characterized by relatively low contextual dependency, we predicted that the enhancing effect of context reexposure on contextual dependency would be larger, whereas the impairing effect of similar context exposure would be smaller, for emotional versus neutral memories.

2. MATERIALS AND METHODS

2.1. Participants.

One hundred eighty subjects (138 females) with a mean age of 21.94 years (*SD*=2.83, range=18-35) participated in the study. We excluded 18 participants who performed at chance level during the recognition test (i.e., in case of a higher than 5 percent chance that recognition responses were random, assessed by way of a binomial test as in previous studies, Zhang et al., 2017). In exchange for participation, the participants received either €24, - (Same Context Exposure group and Similar Context Exposure group) or €16, - (No Exposure group), or an equivalent number of course credit. All procedures were approved by the ethical committee of the Faculty of Social and Behavioral Sciences at Utrecht University.

The participants were allocated to one of six groups. Half of these groups were shown angry faces in a spatial context (Negative Face, final *N*=79), and the other three groups were shown neutral faces (Neutral Face, final *N*=83). On the second day of the experiment, two of the groups were reexposed to the encoding context (Negative Face - Same Context Exposure, *N*=26; Neutral Face - Same Context Exposure, *N*=28), two other groups were exposed to similar contexts as during learning (Negative Face - Similar Context Exposure, *N*=24); Neutral Face - Similar Context Exposure, *N*=28), and the final two groups did not come to the laboratory in between encoding and testing (Negative Face - No Exposure, *N*=29; Neutral Face - No Exposure, *N*=27).

2.2. Stimuli.

2.2.1. Faces. Images of angry (45 male, 45 female) and neutral (45 male, 45 female) faces were drawn from the Radboud Faces Database (Langner et al., 2010), the NimStim set of Facial Expressions (Tottenham et al., 2009), and the Chicago Face Database (Ma et al., 2015) Version 2.0.3., July 2016.

2.2.2. Contexts. As context images, we used a total of 180 pictures of indoor and outdoor spatial environments (e.g., a forest, a kitchen), as in previous research (van Ast et al., 2013, 2014). Two thirds of the images were similar to one of the other images in the set (e.g., two kitchens), and one third of the images were unique.

2.3. Experimental task.

2.3.1. Encoding. An overview of the experimental paradigm is displayed in Fig. 2. During encoding, participants were presented with 30 female and 30 male face images (angry in the Negative Face group, neutral in the Neutral Face group) on unique background pictures. Each

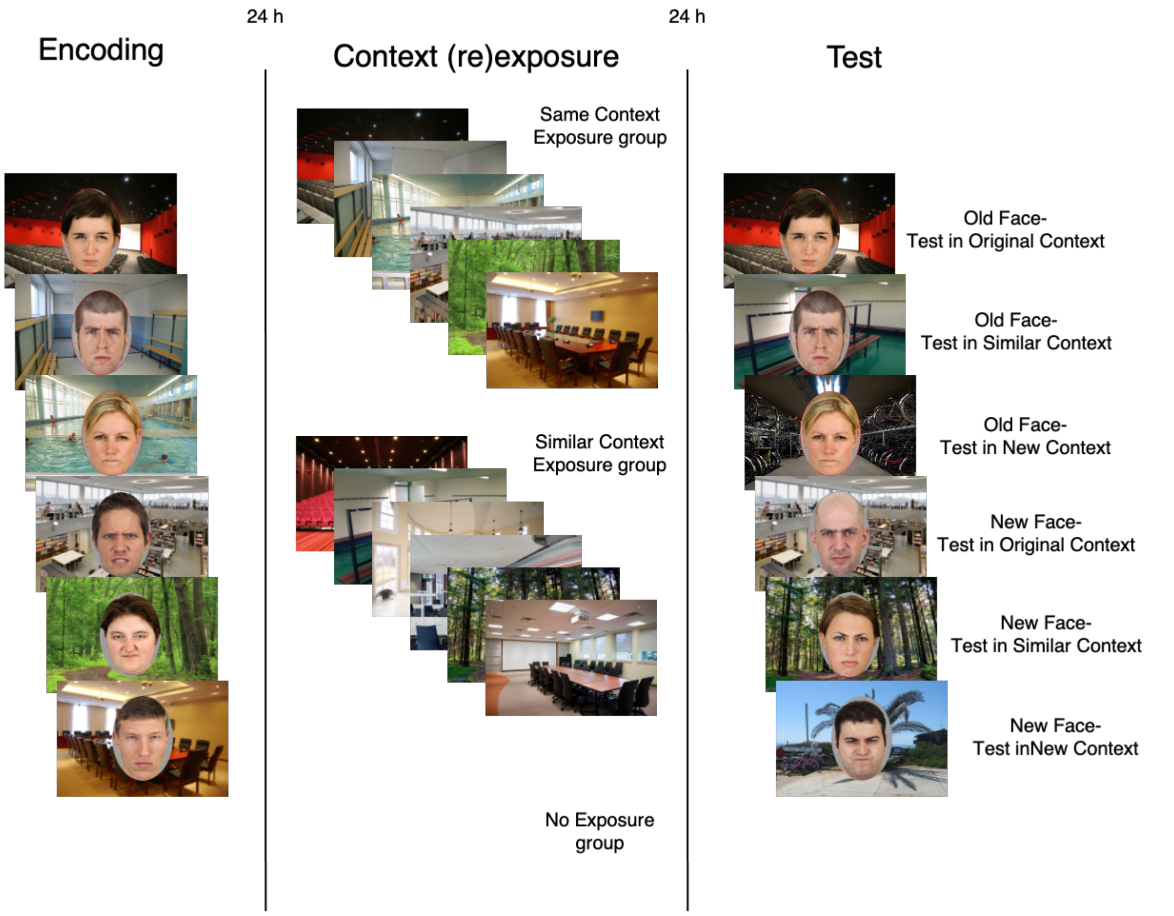

Fig. 2. Experimental paradigm. Participants performed an encoding task, underwent context (re)exposure, and completed a recognition test on three consecutive days. During encoding, participants were randomly assigned to one of two groups and viewed either angry (Negative Face group, displayed here) or neutral faces (Neutral Face group, not displayed), presented on a context image. On day 2, the participants were again randomized, but now to one of three groups. That is, they were either reexposed to all the contexts they had seen on the previous day (Same Context Exposure group), were shown similar contexts to the ones they had seen before (Similar Context Exposure group), or did not come to the laboratory (No Exposure group). On the final day, all participants performed a recognition test, and were presented with faces they had seen on day 1 (Old Face) and unfamiliar faces (New Face). The faces were presented on a context they had seen during encoding (Test in Original Context), a similar context (Test in Similar Context), or a novel context (Test in New Context).

trial started with presentation of the background picture for 2 seconds, after which the face was shown in the middle of the background for 4 seconds. Then the face image disappeared, and only the background was shown for 0.5 more seconds. Next, participants were asked to rate the vividness of their imagined scene on a visual analogue scale, ranging from “not vivid” to “very vivid”. Participants were given 2.5 seconds to respond, during which reaction times were recorded. The next trial started after the presentation of a black background (intertrial interval) for 1, 2, or 3 seconds (randomized in blocks of 3 trials). The participants completed three practice trials before the task commenced.

2.3.2. Context (re)exposure. During the context (re)exposure session, participants were either presented with all 60 background pictures they had seen during encoding (Same Context Exposure group) or shown 60 background picture that resembled the original background picture (Similar Context Exposure group). The No Exposure group did not come to the laboratory. For the Same Context Exposure and Similar Context Exposure groups, the context images were presented – without face images - for 3 seconds, during which participants rated reliving of their previously imagined scenes on a visual analogue scale, ranging from "no reliving" to "much reliving". Reaction times were recorded. The next trial started after the presentation of a black background (intertrial interval) for 1, 2, or 3 seconds. The participants completed three practice trials before the task started.

2.3.3. Test. During the recognition test, participants were shown 30 of the 60 face images they had seen during encoding (i.e., 15 male and 15 female, randomly selected), and 30 new faces (15 male and 15 female, all angry in the Negative Face group and all neutral in the Neutral Face group) as lures. To assess contextual dependency of the memories, 10 of the old face images were presented on the original encoding context of day 1 (Test in Original Context), another 10 on a context similar as the encoding context (Test in Similar Context), and the final 10 on a context that participants had not seen before (Test in New Context). Likewise, the lures were either shown on a context seen on day 1 (images that were not used for recognition of old faces in original contexts), a similar context (images that were not used for recognition of old faces in similar contexts), or a new context (10 new images).

For each trial, the context image was presented first for 3 seconds. Then, the face image appeared during which participants could indicate whether they had seen the face before or not (yes/no). After participants had pressed yes (old) or no (new), they were asked to indicate the level of confidence in their response. If they had responded “no”, they were shown three response options to indicate how sure they were that the face was new (1=sure, 2=probably, 3=guess). When participants had responded “yes”, they were asked to indicate their level of confidence that the face was old (4=guess, 5=probably, 6=sure). Both the yes/no recognition as well as the confidence rating task were self-paced, and response times were recorded. The next trial started after presentation of a black background (intertrial interval) for 1,2, or 3 seconds. Participants performed three practice trial before the task commenced.

2.3.4. Trial order. For each of the three phases (encoding, context (re)exposure, test), the trials were presented in blocks of twelve, in which trials that correspond to each within-subjects condition were equally distributed (e.g., Test in Original Context, Test in Similar Context, Test in New Context for old and new faces). As such, trials from each within-subjects condition were not presented more than three times successively. The order of blocks was the same across the encoding, context (re)exposure, and test phases, but the order of trials within blocks was randomized for each phase.

2.4. Procedure. Upon arrival on day 1, participants were asked to read an information brochure which stated that they would be viewing face images on background pictures, use their imagination during the experiment, and answer questions about the presented images. We purposely kept participants unaware that they would undergo memory tests to prevent them from actively rehearsing the learned material between sessions. After they had given written consent, participants were asked to complete questionnaires. These were collected for potential exploratory purposes, but not further analyzed.

2.4.1. Encoding. Participants were informed that they would be shown background pictures on a computer screen in which a face image would appear after some time. They were instructed to imagine as vividly as possible a scene that involves the face image in the background. It was explained that after the face image and background pictures had disappeared from the screen, they could rate the vividness of their imagined scene within approximately two seconds using a continuous scale.

2.4.2. Context (re)exposure. Twenty-four hours later, participants returned to the lab for a context (re)exposure session. Participants allocated to the Same Context Exposure or Similar Context Exposure groups were explained that they would be shown background images and were instructed to rate within three seconds how vividly they remembered a scene they had imaged on day 1 in relation to the presented image. Participants in the No Exposure group did not come to the lab and thus were not exposed to any of the previously presented contexts.

2.4.3. Test. On the third and final day, participants were asked to complete a recognition test of the faces they had encoded on day 1. It was explained that they would be presented with background pictures and face images, and that some faces had been presented during the encoding session on day 1 while others had not been shown before. They were instructed to respond whether they recognized the face as one they had seen on day 1 (old) or not (new). It was explained that after participants had pressed "old" or "new", they would rate their level confidence in this response (i.e., 1=very sure new, 2=somewhat sure new, or 3=guess new if they had responded "new", and 4=guess old, 5=somewhat sure old, or 6=very sure old if they had responded "old").

After the recognition task, participants were asked to complete a short questionnaire concerning encoding characteristics (e.g., how well they could focus during the task) and overall motivation for each part of the experiment. Finally, they were debriefed and paid for their participation.

2.5. Data analysis.

2.5.1. Vividness.

2.5.1.1. Preprocessing. We calculated average vividness scores for all six groups. We included only 50 of the total 60 vividness trials in our analysis because 10 face-context combinations were not related to any test condition on day 3. That is, Old Face - Test in Original Context, Old Face - Test in Similar Context, Old Face - Test in New Context, New Face - Test in Original Context, and New Face - Test in Similar Context all include a face image or context image that were presented on day 1, which leaves 10 trials that are not relevant for performance on day 3. Participants who did not fill in any vividness rating in time (i.e., within 2.5 seconds) were excluded (Negative Face - Similar Context Exposure, N=1; Negative Face - No Exposure, N=3; Neutral Face - Same Context Exposure, N=2; Neutral Face - No Exposure, N=1).

2.5.1.2. Statistical analyses. To check whether the context (re)exposure groups did not significantly differ in vividness of the imagined scenes on day 1 and explore whether vividness of scenes that included angry versus neutral faces differed, we performed a two-way ANOVA. We tested a main effect of Emotion (Negative Face versus Neutral Face), a main effect of Context (Re)exposure (Same Context Exposure versus Similar Context Exposure versus No Exposure), and an Emotion × Context (Re)exposure interaction. Tukey's HSD tests were performed in case of significant main or interaction effects.

2.5.2. Reliving.

2.5.2.1. Pre-processing. Reliving scores were calculated for the Negative Face - Same Context Exposure, Negative Face - Similar Context Exposure, Neutral Face -Same Context Exposure, and Neutral Face - Similar Context Exposure groups. Again, we included 50 of the 60 trials, as only those were relevant for the memory tests performed on day 3 (see 2.5.1.1.). Participants who did not fill in any reliving rating in time (i.e., within 3 seconds) were excluded (Negative Face - Similar Context Exposure, *N*=1; Neutral Face – Same Context Exposure, *N*=2).

2.5.2.2. Statistical analyses. To assess whether reliving of the scenes that were imaged on day 1 was higher when the same versus a similar context was presented on day 2 and explore whether this effect depends on whether the day 1 memories were emotional, we performed a two-way ANOVA. In this analysis, main effects of Emotion (Negative Face versus Neutral Face), and Context (Re)exposure (Same Context Exposure versus Similar Context Exposure), as well as an Emotion × Context (Re)exposure interaction were included. Tukey's HSD tests were performed in case of a significant interaction effect.

2.5.3. Recognition.

2.5.3.1. Pre-processing. We calculated hit and false alarm rates for all three within-subject conditions (Test in Original Context, Similar Context, or New Context) in all 6 groups. From these hits and false alarms data, we calculated d-prime sensitivity index. To do so, hit rates higher than 0.95, and false alarm rates smaller than 0.05 were truncated at 0.95 and 0.05, respectively (Stanislaw & Todorov, 1999).

Subsequently, we calculated two difference scores (Test in Original Context – Test in New Context, and Test in Similar Context – Test in New Context). These constitute the variables of interest to assess changes in contextual dependency of memory (Contextualization Original Context and Contextualization Similar Context, respectively).

2.5.3.2. Statistical analyses.

2.5.3.2.1. Manipulation check: Contextual dependency of emotional and neutral memories. To assess whether correct recognition depended on context, and this contextual dependency was smaller for emotional than neutral memories, we analyzed d-prime scores in the No Exposure control group. We only included the control group for this analysis, because in earlier studies that we aimed to replicate here no exposure to context was included between encoding and testing (Cox, Meeter, et al., 2022; van Ast et al., 2013, 2014). To test if contextual dependency of emotional memories is impaired relative to neutral memories, we performed two independent t-tests. The independent variable was Emotion (Negative Face versus Neutral Face) and dependent variables were d-prime difference scores (Contextualization Original Context, and Contextualization Similar Context). If contextual dependency of emotional memory is relatively low, then the Contextualization Original Context and Contextualization Similar Context scores should *both* be smaller in the Negative Face than the Neutral Face group (see 2.5.3.1.). If these effects were indeed observed, we analyzed contextual dependency of emotional and neutral memories separately by way of repeated measures ANOVAs with Test (Test in Original Context, Test in Similar Context, Test in New Context) as the independent variable, and the d-prime score as the dependent variable. Planned comparisons were performed to compare accuracies between the conditions if an effect of Test was observed. When no relatively low contextual dependency of emotional memory was found, we collapsed the data of the Negative Face and Neutral Face conditions.

2.5.3.2.2. Hypothesis test: Effects of context (re)exposure on contextual dependency. To assess whether context reexposure modulates contextual dependency of memory, we performed univariate ANOVAs with Emotion (Negative Face, Neutral Face) and Context (Re)exposure (Same Context Exposure, Similar Context Exposure, No Exposure) as independent variable and the d-prime difference scores (Contextualization Original Context, and Contextualization Similar Context) as dependent variables. If no Emotion × Context (Re)exposure interaction was found, we collapsed the data of the Negative Face and Neutral Face groups to test for an effect of Context (Re)exposure on contextual dependency. In case of significant effects, we performed planned contrasts to see if (i) the Contextualization Original Context score is higher, but the Contextualization Similar Context is lower in the Same Context Exposure group versus the No Exposure group (i.e., greater contextual dependency, see 2.5.3.1), whereas (ii) both the Contextualization Original Context and Contextualization Similar Context are lower in the Similar Context Exposure group versus the No Exposure group (i.e., smaller contextual dependency).

2.5.4. Interactions between reliving strength and changes in contextual dependency by context (re)exposure.

2.5.4.1. Preprocessing. To regress possible changes in contextual dependency on retrieval strength, we linked the recognition responses for old and new faces on day 3 with the respective reliving scores during context exposure on day 2 (0-100) on the level of trials for 5 within-subject conditions (Old Face - Test in Original, Similar Context, or New Context, New Face - Test in Original or Similar Context). Note that the new faces (lures) were of course themselves never presented during context (re)exposure, but the contexts they were presented on during testing were in fact shown in this phase. However, new faces presented on new contexts on day 3 are (by definition) not related to any reliving scores on day 2, such that no corresponding reliving score exists for New Face - Test in New Context. This absence of a relationship between recognition response and retrieval strength needs to be included in the model to predict d-prime accuracy in new contexts that are based on recognition of old faces. Since 10 of the 60 trials on day 2 (context (re)exposure) are not related to any test trial on day 3 (see sections 2.5.1.1. and 2.5.2.1.), these reliving trials can be used to create this absence of a relationship between recognition responses and reliving for New Face - Test in New Context. Therefore, we randomly paired the reliving scores of these 10 remaining trials on day 2 with recognition responses during presentation of new faces on new contexts on day 3. Due to this random pairing, no relationship should emerge such that the Test in New Context condition can be included in the model as reference condition.

2.5.4.2. Statistical analyses. To assess if contextual dependency (i) becomes larger in the Same Context Exposure group, but (ii) smaller in the Similar Context Exposure group, with stronger memory retrieval during context (re)exposure, we performed a multilevel probit regression. The trial level coefficients yielded by this type of model are Z-scores and have been shown to be mathematically equivalent to d-prime (Decarlo, 1998), the key measure from signal detection theory that is typically computed by aggregating data within participants. Therefore, their interpretation is the same as d-prime: the beta parameters of this regression represent an increase or decrease in the probability of a 'yes' response associated with each variable. As fixed effects on the item level (1^{st} level) we entered Reliving (0-100, group-mean centered), Face (Old, New), and Test (Test in Original Context, Test in Similar Context, Test in New Context). On the participant level (2^{nd} level) we entered Emotion (Negative Face, Neutral Face). The Same Context Exposure and Similar Context Exposure groups were analyzed separately. Random intercepts were included for participants (Twisk, 2006).

First, we aimed to assess whether the random pairing of reliving scores with recognition responses for New Face - Test in New Context was successful and thus appropriate to use as reference (see 2.5.4.1.): here no significant relationship should occur between reliving strength and memory accuracy. For the role of reliving strength during context (re)exposure in later contextual dependency of memory, we expected that higher reliving in the Same Context Exposure group would predict (i) improved memory for Test in Original Context versus Test in New Context (i.e., Contextualization Original Context) and (ii) reduced memory for Test in Similar Context versus Test in New Context (i.e., Contextualization Similar Context). Furthermore, we expected that higher reliving in the Similar Context Exposure group would predict (iii) reduced memory for both Test in Original Context versus Test in New Context (i.e., Contextualization Original Context) and (iv) Test in Similar Context versus Test in New Context (i.e., Contextualization Similar Context). We first tested an interaction with Emotion for each of these effects (i.e., Emotion × Reliving × Test × Face). If no significant 4-way interactions were observed, we collapsed the data of the Negative Face and Neutral Face groups. We then tested Reliving × Test (Test in Original Context versus Test in New Context) × Face (Old versus New) and Reliving × Test (Test in Similar Context versus Test in New Context) × Face (Old versus New) interactions for the Same Context Exposure group and the Similar Context Exposure group separately. Planned comparisons were performed in case of a significant effect to assess the predicted relationships between reliving strength and the test conditions. Alpha was set at .05 for all tests.

3. RESULTS

3.1. Manipulation checks.

3.1.1. Vividness. Vividness of the imagined scenes during encoding was lower for negative than neutral memories (Emotion, $F_{1,150} = 7.08$, $p = .009$, Fig. 3), in line with earlier research (de Vries et al., 2022). Crucially, there were no significant differences in vividness between the Same Context Exposure, Similar Context Exposure, and No Exposure groups (Context (Re)exposure, $F_{2,150} = 0.14$, $p = .869$). We also did not find that whether memories were negative or neutral modulated this effect (Emotion × Context (Re)exposure, $F_{2,150} = 1.44$, $p = .241$). These findings confirm that no unexpected differences in vividness occurred between the groups.

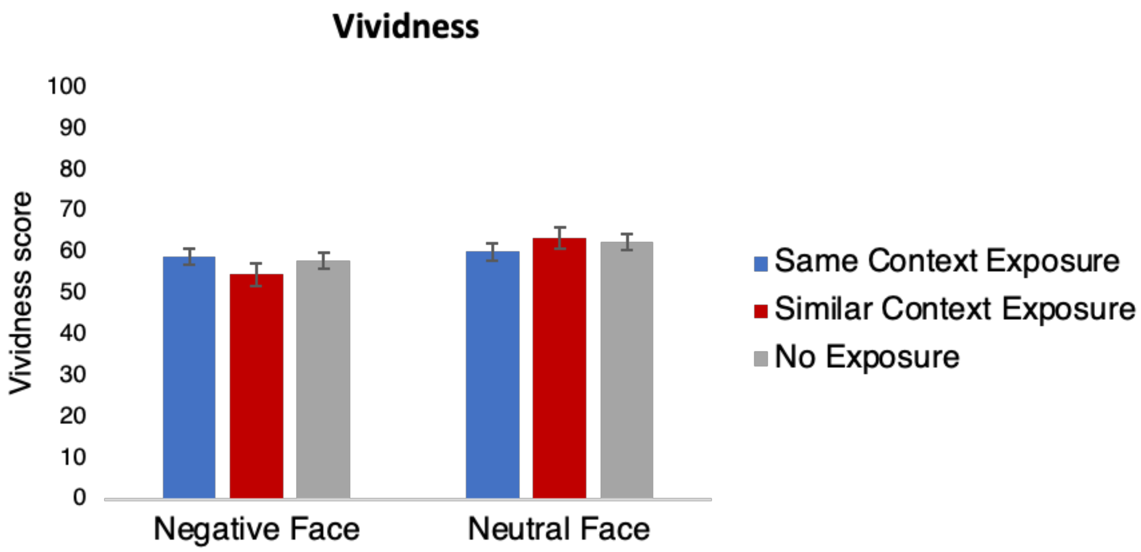

Fig. 3. Vividness ratings during the encoding task in the Same Context Exposure (blue bars), Similar Context Exposure (red bars), and No Exposure group (grey bars). The scores are plotted separately for the Negative Face (left) and Neutral Face (right) groups. Error bars represent SEM.

3.1.2. Reliving. Analyses of reliving ratings on day 2 confirmed that context reexposure led to more reliving than exposure to a similar context (Context (Re)exposure, $F_{1,99} = 97.01$, $p < .001$, Fig. 4). This effect did not depend on whether memories were negative or neutral (Emotion × Context (Re)exposure, $F_{1,99} = 0.3$, $p = .720$). Finally, there was no overall difference in reliving for negative versus neutral memories (Emotion, $F_{1,99} = 0.52$, $p = .474$). In short, inducing relatively strong reliving by reexposure to the encoding context was successful, and no other unexpected effects on reliving were observed.

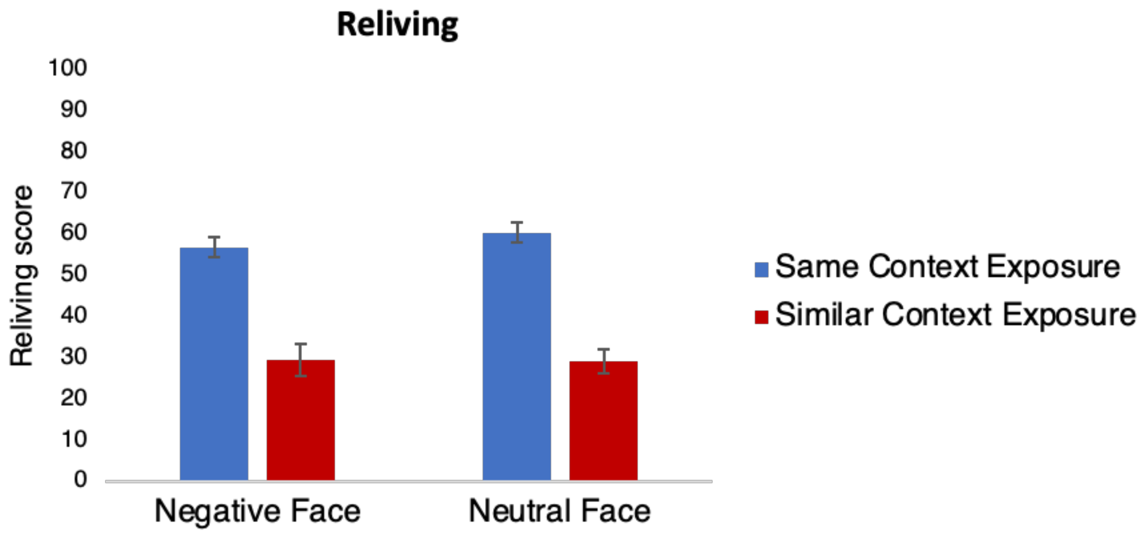

Fig. 4. Reliving ratings in the Same Context Exposure (blue bars) and Similar Context Exposure (red bars) group. The scores are plotted separately for the Negative Face (left) and Neutral Face (right) groups. Error bars represent SEM.

3.1.3. Contextual dependency of emotional and neutral memories. We observed that, for the No Exposure groups, the Contextualization Original Context score ($t_{54} = 1.78$, $p = .081$) and the Contextualization Similar Context score ($t_{54} = 2.09$, $p = .041$) were at least trend significantly smaller in the Negative Face group than the Neutral Face group (Fig. S1). These findings thus suggest that contextual dependency was indeed lower for emotional than neutral memories. Furthermore, for neutral memories, recognition accuracy was dependent on context (Test, $F_{2,52} = 9.83$, $p < .001$, Fig. S2B, right pane). Planned comparisons suggested that accuracy scores were highest in the original encoding context, lower in the similar context, and lowest in the new context (Test in Original Context versus Test in Similar Context, $F_{1,26} = 6.61$, $p = .016$; Test in Similar Context versus Test in New Context, $F_{1,26} = 3.65$, $p = .067$). Emotional memories also depended on context (Test, $F_{2,56} = 3.67$, $p = .002$, Fig. S2A, right pane), but here recognition was only elevated in the original versus similar context (Test in Original Context versus Test in Similar Context, $F_{1,28} = 8.50$, $p = .007$; $F < 2.34$, $p > 0.14$ for other comparisons). In short, these results largely replicate earlier studies showing context-dependent memory (Cox, Meeter, et al., 2022; Godden & Baddeley, 1975;

Smith & Vela, 2001; van Ast et al., 2013, 2014), and smaller contextual dependency for emotional versus neutral memories (Cox, Meeter, et al., 2022; van Ast et al., 2013, 2014).

3.2. Hypothesis tests.

3.2.1. Effects of context (re)exposure on contextual dependency. Whether memories were emotional or neutral did not significantly influence any effects of context (re)exposure on contextual dependency (Emotion × Context (Re)exposure, $F_{2,156} = 0.38$, $p = .687$ for Contextualization Original Context; Emotion × Context (Re)exposure, $F_{2,156} = 2.84$, $p = .061$ for Contextualization Similar Context). Therefore, we collapsed the data of the Negative Face and Neutral Face groups. Crucially, contextual dependency did not differ between the context (re)exposure groups in the collapsed data ($F_{2,156} = 1.46$, $p = .236$ for Contextualization Original Context and $F_{2,156} = 1.80$, $p = .169$ for Contextualization Similar Context, Fig. 5 and Fig. S2-S4 for more detail).

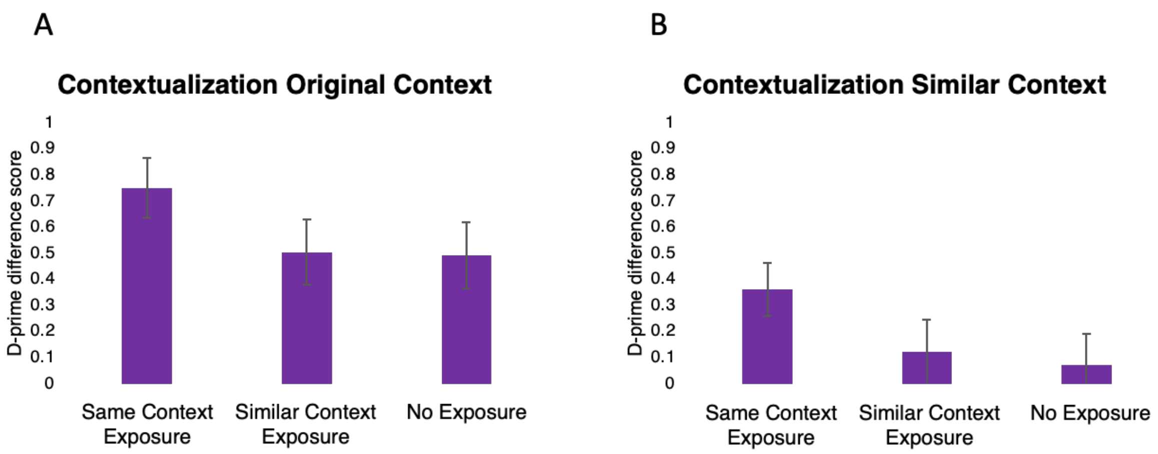

Fig. 5. (**A**) Contextualization Original Context and (**B**) Contextualization Similar Context scores in the Same Context Exposure (left), Similar Context Exposure (middle), and No Exposure (right) groups (collapsed over the Negative Face and Neutral Face groups). Error bars represent SEM.

3.2.2. Interactions between reliving strength and changes in contextual dependency by context (re)exposure. We first assessed whether the impact of reliving in the reference condition was indeed absent. That is, we tested whether including new faces presented in new contexts in the model related to reliving of the randomly paired contexts. Here, we found no main effect of reliving nor an interaction between Reliving and Emotion in neither the Same Context Exposure nor the Similar Context Exposure group (Same Context Exposure group: Reliving, $Z = -0.05$, $p = .623$, Reliving × Emotion (Negative versus Neutral), $Z = 0.02$, $p = .895$; Similar Context Exposure group: Reliving, $Z = 0.02$, $p = .786$, Reliving × Emotion (Negative versus Neutral), $Z = -0.09$, $p = .527$). This shows that random pairing of reliving scores with recognition responses was successful and Test in New Context can be used as the reference condition in the models (see 2.5.4.1 and 2.5.4.2). Whether memories were negative or neutral did not modulate any interactions between reliving strength and contextual dependency (Same Context Exposure group: Emotion (Negative versus Neutral) × Reliving × Test (Test in Original Context versus Test in New Context) × Face (Old versus New), $Z = 0.12$, $p = .639$, Emotion (Negative versus Neutral) × Reliving × Test (Test in Similar Context versus Test in New Context) × Face (Old versus New), $Z = -0.23$, $p = .366$; Similar Context Exposure group: Emotion (Negative versus Neutral) × Reliving × Test (Test in Original Context versus Test in New Context) × Face (Old versus New), $Z = -0.06$, $p = .810$, Emotion (Negative versus Neutral) × Reliving × Test (Test in Similar Context versus Test in New Context) × Face (Old versus New), $Z = 0.21$, $p = .425$). Therefore, we again collapsed the data of the Negative Face and Neutral Face groups. Crucially, and in contrast with the predictions, reliving did not relate to the strength of contextual dependency in either the Same Context Exposure group (Reliving × Test (Test in Original Context versus Test in New Context) × Face (Old versus New), $Z = -0.05$, $p = .701$; Reliving × Test (Test in Similar Context versus Test in New Context) × Face (Old versus New), $Z = -0.05$, $p = .700$), nor the Similar Context Exposure group (Reliving × Test in Original Context versus Test in New Context) × Face (Old versus New), $Z = 0.00$, $p = .979$;

Reliving × Test (Test in Similar Context versus Test in New Context) × Face (Old versus New), $Z = 0.09$, $p = .510$).

4. DISCUSSION

In the present study, we aimed to test whether contextual binding of human episodic memory can be changed by (re)exposure to contexts that are perceptually identical or resemble the original encoding context. We also assessed whether these effects are predicted by the strength of memory retrieval during (re)exposure and differ in magnitude for emotional versus neutral memories. We found that correct recognition was dependent on test context and that this contextual dependency was lower for emotional than neutral memories. We also observed that memory retrieval was stronger during reexposure to the original encoding context than during exposure to a similar context. These findings replicate earlier studies (Cox, Meeter, et al., 2022; van Ast et al., 2013, 2014), and show that our basic manipulations were successful. However, we did not find evidence for any of the hypotheses: reexposure to the original encoding context did not promote contextual dependency of memory, nor did exposure to a similar context reduce it. Also, no relationships were observed between retrieval strength during (re)exposure to the contexts and contextual dependency at test. Finally, no differences were found between emotional and neutral memories in any of these hypothesized effects. Therefore, the findings of this large-scale study do not support the idea that context (re)exposure after a neutral or emotional event modulates the extent to which subsequent memory activation becomes restricted to contextual cues that were present during encoding.

These observations clearly contrast with fear-conditioning studies that did show effects of context (re)exposure on fear generalization in animals (Al Abed et al., 2020; de Oliveira Alvares et al., 2012; Sekeres et al., 2020; Sevenster et al., 2018; Wiltgen & Silva, 2007; Winocur et al., 2009; Zhou & Riccio, 1994) and humans (Sevenster et al., 2017). One obvious difference between these previous studies and the present experiment is the emotional intensity of the memories. Whereas fear conditioning involves the administration of – sometimes severe - electric shocks, we here presented threatening faces, which in all likelihood are experienced as only mildly fearsome. It has been suggested that impaired contextual processing of emotional memories, by amygdalar downregulation of the hippocampus, may only occur under particularly strong conditioning protocols (Layton et al., 2002), which are more aversive than the procedures adopted here. One could thus suspect that the present design was not as well suited to investigate changes in contextual dependency of emotional memories as fear-conditioning paradigms. However, it is important to note that we did observe the expected smaller contextual dependency of emotional versus neutral memories, like in previous studies (Cox, Meeter, et al., 2022; van Ast et al., 2013, 2014). Apparently, the threatening faces were arousing enough to at least partly impair the processing of context, such that the premise to study changes in contextual binding of emotional events was met.

One other major difference between previous studies (Al Abed et al., 2020; de Oliveira Alvares et al., 2012; Sekeres et al., 2020; Sevenster et al., 2017, 2018; Wiltgen & Silva, 2007; Winocur et al., 2009; Zhou & Riccio, 1994) and the present experiment is the outcome variable: we tested for contextual dependency of declarative memory instead of conditioned fear responses. Having the possibility to check whether contextual dependency was smaller for emotional than neutral memories was one of the reasons we adopted this episodic memory paradigm. This method also allowed for the inclusion of many trials per participant, which enabled us to assess relationships between memory reliving during context (re)exposure and contextual dependency at test in a statistically powerful manner. Even though it is not known how episodic memory contextualization maps onto fear generalization, it has been proposed that contextual processing deficits may subserve each (e.g., Maren et al., 2013). Likewise, broad impairment in hippocampus-dependent associative learning have been revealed as a vulnerability factor for PTSD (Lambert & McLaughlin, 2019) and associative learning of foreground cues and background context has been proposed as an essential ingredient to form an integrated representation of an event (Lambert & McLaughlin, 2019). Thus, at least in theory, one could suspect that contextual dependency of declarative memory and conditioned fear responses could be affected by context (re)exposure in similar ways.

Nevertheless, it must be noted that there are several disadvantages to the paradigm we used here, which may have prevented us from observing the hypothesized effects of context (re)exposure. In animal fear-conditioning research, rodents typically explore a context for several minutes during fear conditioning (e.g., Sekeres et al., 2020). The extensive exploration of the context, together with the occurrence of a threatening shock, leads to the formation of a robust contextual memory. This firmly established memory can then be updated by reexploration, through the integration of contextual information that the animal had not yet learned or had already forgotten (Al Abed et al., 2020). In the paradigm adopted here, initial learning involved the imagination of many scenes based on faces of similar expressions (either angry or neutral) in contexts, which were presented in quick succession. Conceivably, this method made it somewhat difficult for participants to imagine clear and well-defined episodes and therefore did not lead to the formation of complete episodic memories. Without clearly established memories, it is possible that during context (re)exposure participants did not learn information to integrate with the existing memory. In future research, it may be important to use a procedure that promotes more unique associations between learning events and their spatial context, and to induce life-like episodic memories (Tulving, 1972) that consist of what-where-when qualities and a recollective experience during recall (Tulving, 2002). We have recently shown that this can be accomplished by using words for central events (Cox, Dobbelaar, et al., 2021). Presenting images of objects could also be a more effective way to induce unique episodic events (de Vries et al., 2022). Apart from difficulties in the formation of complete episodic memories during learning, the relatively little time to encode new or lost contextual information during context (re)exposure could also have been a limiting factor on integration. Finally, participants were perhaps not motivated enough to learn during context (re)exposure, because associations between context images and angry faces are not as threatening (and therefore relevant) to the participants like context-shock pairings are to animals in a fear-conditioning paradigm. Thus, even though we confirmed that emotional memories were less dependent on context than neutral memories, the present paradigm may have obstructed in several ways the integration of new information with existing memory such that contextual dependency remained unaffected by our manipulations.

In conclusion, our findings suggest that manipulating contextual dependency of human episodic memory by exposure to the same or a similar context as during encoding is challenging. So far, the mechanisms underlying the use of context in therapy have not been formalized, nor have these been put to the test. Here, we have performed a step in this direction. Revisiting the site of a traumatic event has been described as an effective component of psychotherapy (Ehlers & Clark, 2009) and animal research shows promising effects of context reexposure on fear generalization (Al Abed et al., 2020; de Oliveira Alvares et al., 2012; Sekeres et al., 2020; Sevenster et al., 2017, 2018; Wiltgen & Silva, 2007; Winocur et al., 2009; Zhou & Riccio, 1994). Nevertheless, clinical approaches typically aim to directly target the emotional hotspots of memory (de Voogd et al., 2018; Holmes et al., 2005; Mineka & Oehlberg, 2008), thereby potentially overlooking a complementary route towards desirable treatment outcomes. A "recontextualization intervention" to release patients from dysfunctional activation of emotional memories is currently not within direct reach. Research aimed at revealing the conditions under which memory integration operates seems key to change this. Such fundamental insight may improve our understanding of how contextual dependency of memory can be targeted and guide us towards successful ways to keep emotional memories at bay.

Author Contributions

Wouter Cox: Formal Analysis, Data Curation, Writing- Original Draft Preparation, Visualization. **Mandy Woelk**: Investigation, Data Curation, Writing- Reviewing and Editing. **Olivier de Vries**: Formal Analysis, Writing- Reviewing and Editing. **Angelos Krypotos**: Software, Writing- Reviewing and Editing. **Merel Kindt**: Writing- Reviewing and Editing. **Iris Engelhard**: Writing- Reviewing and Editing, Funding Acquisition. **Dieuwke Sevenster**: Conceptualization, Methodology, Supervision, Project Administration. **Vanessa van Ast**: Conceptualization, Methodology, Resources, Writing- Reviewing and Editing, Supervision, Project Administration.

Funding

This work was supported by a NWO Vici Grant (Iris Engelhard, 453-15-005). Wouter Cox is supported by a NWO Research Talent Grant (406-16-557), Merel Kindt by a ERC Advanced Grant (743263), and Vanessa van Ast by a NWO Veni Grant (451-16-021).

Literature

Acheson, D. T., Gresack, J. E., & Risbrough, V. B. (2012). Hippocampal dysfunction effects on context memory: Possible etiology for posttraumatic stress disorder. *Neuropharmacology*, *62*, 674–685.

Al Abed, A. S., Ducourneau, E. G., Bouarab, C., Sellami, A., Marighetto, A., & Desmedt, A. (2020). Preventing and treating PTSD-like memory by trauma contextualization. *Nature Communications*, *11*, 4220.

Bisby, J. A., King, J. A., Brewin, C. R., Burgess, N., & Curran, H. V. (2010). Acute effects of alcohol on intrusive memory development and viewpoint dependence in spatial memory support a dual representation model. *Biological Psychiatry*, *68*, 280–286.

Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychological Review*, *117*, 210–232.

Cox, W. R., Dobbelaar, S., Meeter, M., Kindt, M., & van Ast, V. A. (2021). Episodic memory enhancement versus impairment is determined by contextual similarity across events. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2101509118.

Cox, W. R., Meeter, M., Kindt, M., & van Ast, V. A. (2022). Time-dependent emotional memory transformation: Divergent pathways of item memory and contextual dependency. *Journal of Experimental Psychology: General*. Advance online publication.

de Oliveira Alvares, L., Einarsson, E. O., Santana, F., Crestani, A. P., Haubrich, J., Cassini, L. F., Nader, K., & Quillfeldt, J. A. (2012). Periodically reactivated context memory retains its precision and dependence on the hippocampus. *Hippocampus*, *22*, 1092–1095.

de Voogd, X. L. D., Kanen, J. W., Neville, X. D. A., Roelofs, X. K., Fernandez, G., & Hermans, E. J. (2018). Eye-movement intervention enhances extinction via amygdala deactivation. *Journal of Neuroscience*, *38*, 8694–8706.

de Vries, O. T., Grasman, R. P. P. P., Kindt, M., & Ast, V. A. van. (2022). Threat learning impairs subsequent associative inference. *Scientific Reports, 12,* 18878.

Decarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological*

Methods, *3*, 186–205.

Desmedt, A., Marighetto, A., & Piazza, P. V. (2015). Abnormal fear memory as a model for posttraumatic stress disorder. *Biological Psychiatry*, *78*, 290–297.

Dimsdale-Zucker, H., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal sub fi elds. *Nature Communications*, *9*, 294.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*, 319–345.

Ehlers, A., Clark, D. M., Ehlers, A., & Clark, D. M. (2009). Post-traumatic stress disorder: The development of effective psychological treatments *Nordic Journal of Psychiatry*, *62*, 11–18.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*, 109–120.

Foa, E. B., Dancu, C. V., Hembree, E. A., Jaycox, L. H., Meadows, E. A., & Street, G. P. (1999). A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. *Journal of Consulting and Clinical Psychology*, *67*, 194–200.

Fujinaka, A., Li, R., Hayashi, M., Kumar, D., Changarathil, G., Naito, K., Miki, K., Nishiyama, T., Lazarus, M., Sakurai, T., Kee, N., Nakajima, S., Wang, S. H., & Sakaguchi, M. (2016). Effect of context exposure after fear learning on memory generalization in mice. *Molecular Brain*, *9*, 3–9.

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

Holmes, E. A., Arntz, A., & Smucker, M. R. (2007). Imagery rescripting in cognitive behaviour therapy: Images , treatment techniques and outcomes. *Journal of Behavior*

Therapy and Experimental Psychiatry, *38*, 297–305.

Holmes, E. A., Grey, N., & Young, K. A. D. (2005). Intrusive images and “hotspots” of trauma memories in posttraumatic stress disorder: an exploratory investigation of emotions and cognitive themes. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*, 3–17.

Josselyn, S. A., & Frankland, P. W. (2018). Memory allocation: Mechanisms and function. *Annual Review of Neuroscience*, *41*, 389–413.

Kunze, A. E., Lancee, J., Morina, N., Kindt, M., & Arntz, A. (2019). Mediators of change in imagery rescripting and imaginal exposure for nightmares: Evidence from a randomized wait-list controlled trial. *Behavior Therapy*, *50*, 978–993.

Lambert, H. K., & Mclaughlin, K. A. (2019). Impaired hippocampus-dependent associative learning as a mechanism underlying PTSD: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *107*, 729–749.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, *24*, 1377–1388.

Layton, B., Ph, D., & Krikorian, R. (2002). Memory mechanisms in posttraumatic stress disorder. *Journal Of Neuropsychiatry*, *14*, 254–261.

Liberzon, I., & Abelson, J. L. (2016). Context processing and the neurobiology of post-traumatic stress disorder. *Neuron*, *92*, 14–30.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122–1135.

Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, *14*, 417–428.

Meyer, T., Krans, J., van Ast, V., & Smeets, T. (2017). Visuospatial context learning and configuration learning is associated with analogue traumatic intrusions. *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 120–127.

Mineka, S., & Oehlberg, K. (2008). The relevance of recent developments in classical conditioning to understanding the etiology and maintenance of anxiety disorders. *Acta Psychologica*, *127*, 567–580.

Sekeres, M. J., Moscovitch, M., Grady, C. L., Sullens, D. G., & Winocur, G. (2020). Reminders reinstate context-specificity to generalized remote memories in rats: Relation to activity in the hippocampus and aCC. *Learning & Memory*, *27*, 1–5.

Sevenster, D., de Oliveira Alvares, L., & D'Hooge, R. (2018). Pre-exposure and retrieval effects on generalization of contextual fear. *Learning and Motivation*, *63*, 20–26.

Sevenster, D., Haesen, K., Vervliet, B., Kindt, M., & D'Hooge, R. (2017). Prevention and treatment strategies for contextual overgeneralization. *Scientific Reports*, *7*, 1–14.

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

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *3*, 137–149.

Tottenham, N., Tanaka, J. W., Leon, A. C., Mccarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 382–402). Academic Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25.

van Ast, V. A., Cornelisse, S., Meeter, M., Joëls, M., & Kindt, M. (2013). Time-dependent effects of cortisol on the contextualization of emotional memories. *Biological Psychiatry*, *74*, 809–816.

van Ast, V. A., Cornelisse, S., Meeter, M., & Kindt, M. (2014). Cortisol mediates the effects of stress on the contextual dependency of memories. *Psychoneuroendocrinology*, *41*, 97–110.

van Kesteren, M. T. R., Krabbendam, L., & Meeter, M. (2018). Integrating educational knowledge: Reactivation of prior knowledge during educational learning enhances memory integration. *Npj Science of Learning*, *3*, 11.

van Kesteren, M. T. R., Rignanese, P., Gianferrara, P. G., Krabbendam, L., & Meeter, M. (2020). Congruency and reactivation aid memory integration through reinstatement of prior knowledge. *Scientific Reports*, *10*, 4776.

Voorendonk, E. M., Meyer, T., Duken, S. B., & van Ast, V. A. (2021). Cardiorespiratory fitness as protection against the development of memory intrusions: A prospective trauma analogue study. *Biological Psychology*, *165*, 108189.

Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning and Memory*, *14*, 313–317.

Winocur, G., Frankland, P. W., Sekeres, M., Fogel, S., & Moscovitch, M. (2009). Changes in context-specificity during memory reconsolidation: Selective effects of hippocampal lesions. *Learning and Memory*, *16*, 722–729.

Wright, D. B., Ruth, H., & M, S. E. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*, *41*, 257–267.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in*

Neurosciences, *34*, 515–525.

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*, 168–179.

Zhang, W., van Ast, V. A., Klumpers, F., Roelofs, K., & Hermans, E. J. (2017). Memory contextualization: The role of prefrontal cortex in functional integration across item and context representational regions. *Journal of Cognitive Neuroscience*, *30*, 579–593.

Zhou, Y. L., & Riccio, D. C. (1994). Pretest cuing can alleviate the forgetting of contextual stimulus attributes. *Learning and Motivation, 25,* 233–244.

SUPPLEMENTARY FIGURES

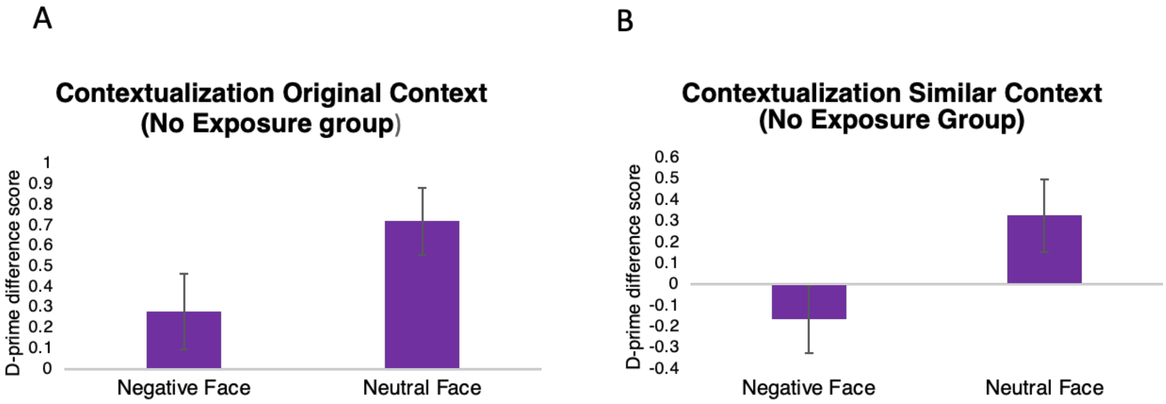

Fig. S1. (**A**) D-prime Contextualization Original Context (left) and (**B**) Contextualization Similar Context scores (right) in the No Exposure groups (separately plotted for the Negative Face and Neutral Face groups). Error bars represent SEM.

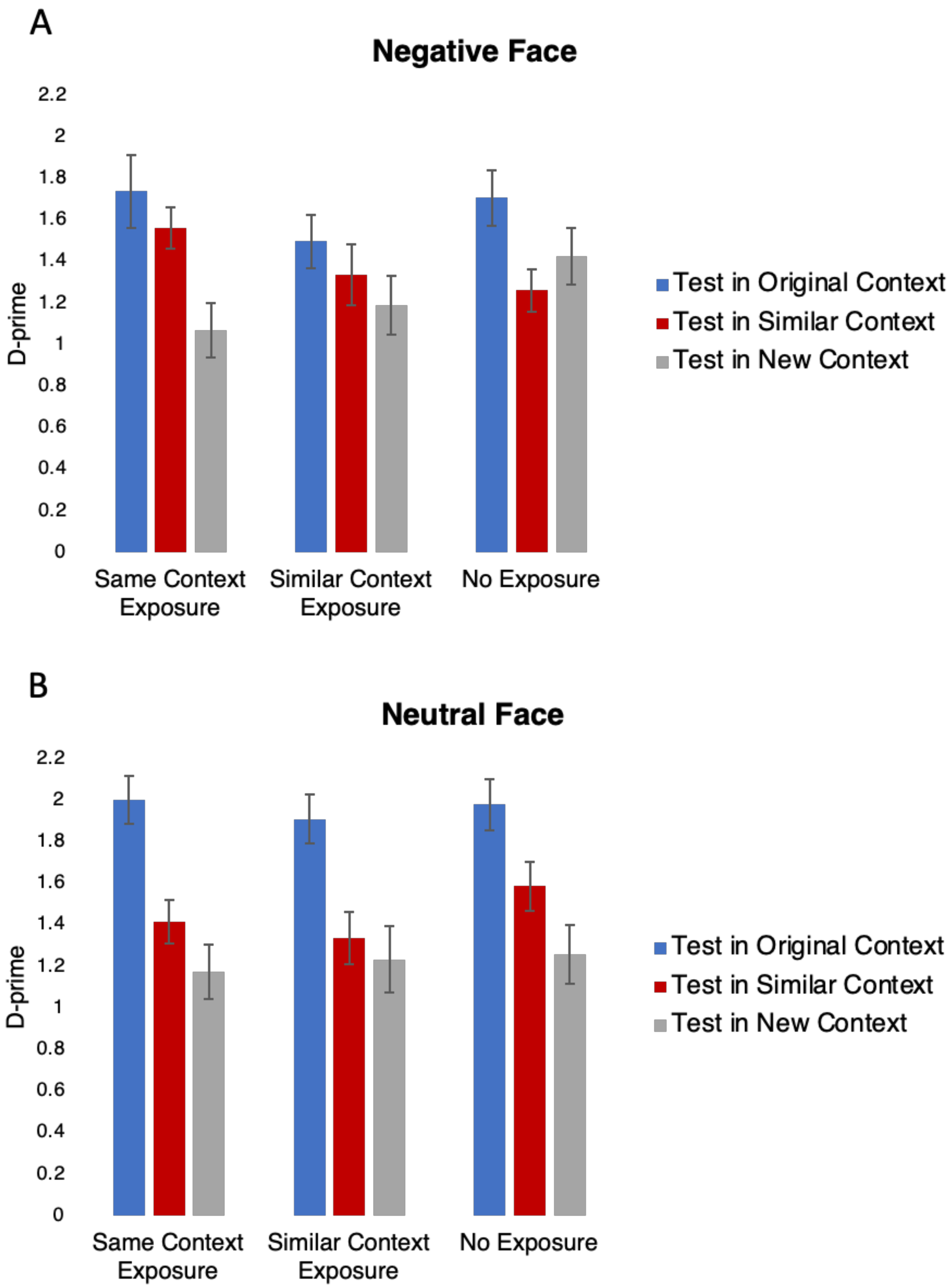

Fig. S2. D-prime accuracy scores in the Negative Face (**A**) and Neutral Face (**B**) groups for the Test in Original Context (blue), Test in Similar Context (red), and Test in New Context (grey) conditions. The scores are plotted separately for the Same Context Exposure (left), Similar Context Exposure (middle), and No Exposure (right) groups. Error bars represent SEM.

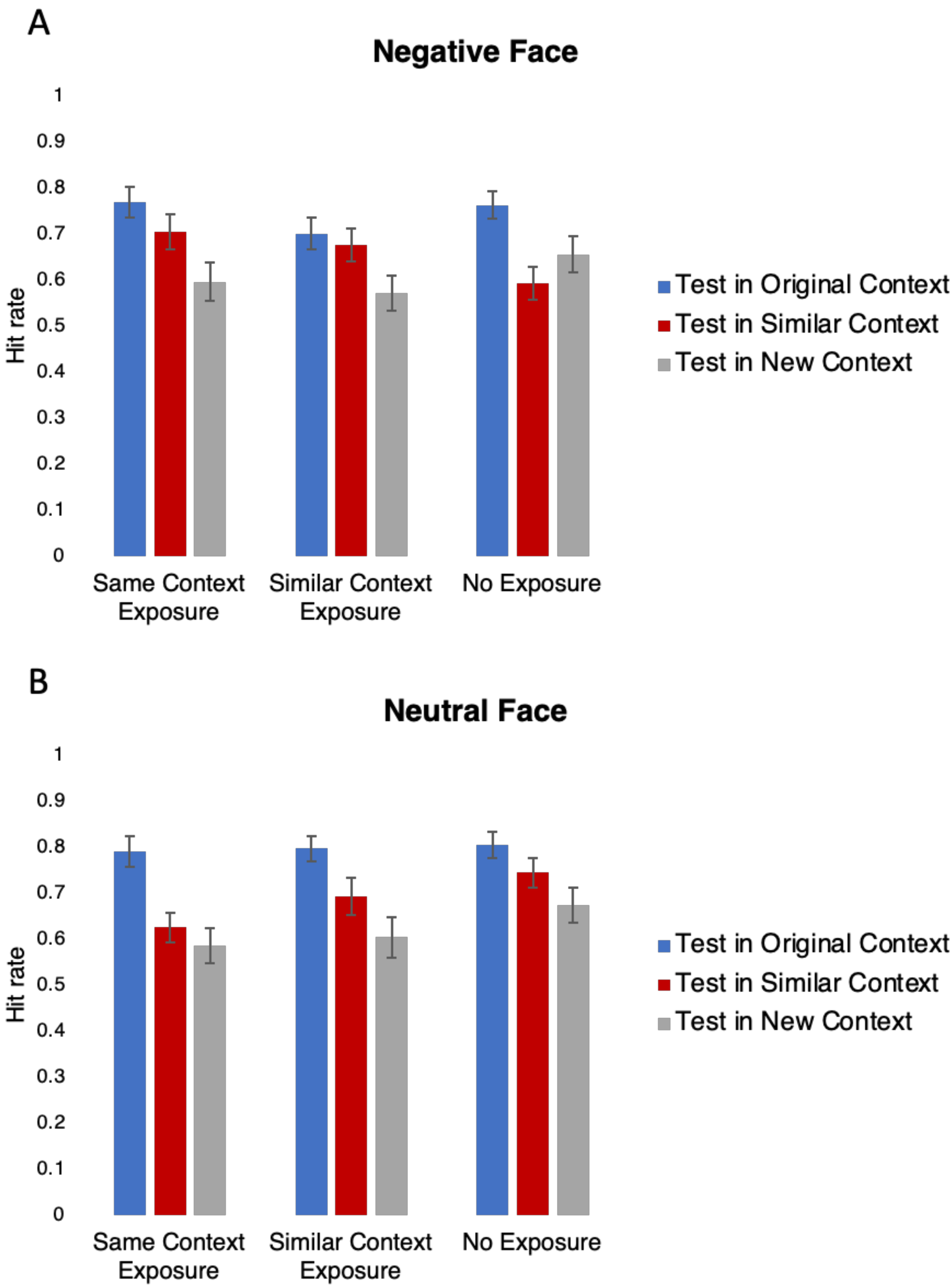

Fig. S3. Hit rates in the Negative Face (**A**) and Neutral Face (**B**) groups for the Test in Original Context (blue), Test in Similar Context (red), and Test in New Context (grey) conditions. The scores are plotted separately for the Same Context Exposure (left), Similar Context Exposure (middle), and No Exposure (right) groups. Error bars represent SEM.

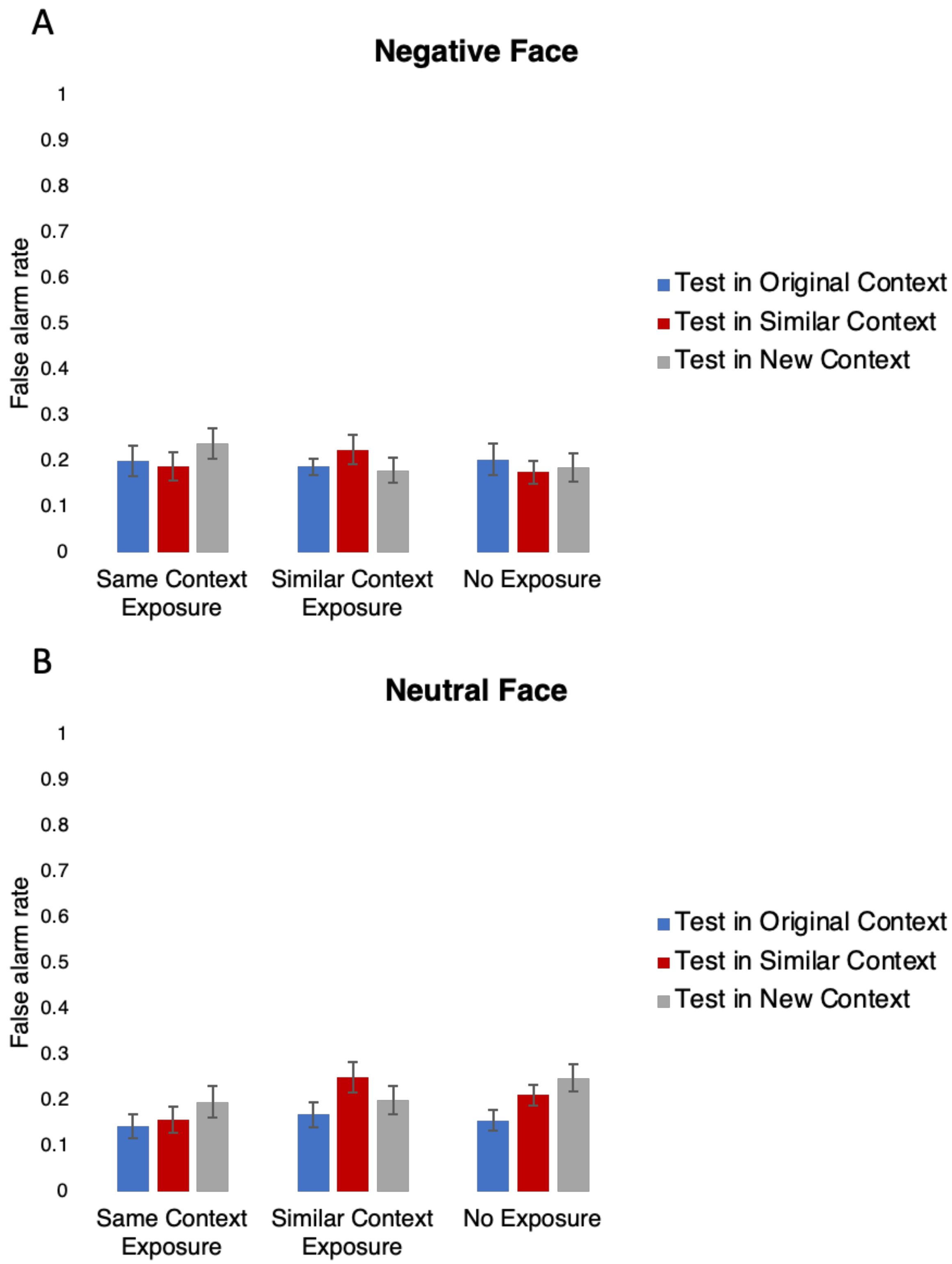

Fig. S4. False alarm rates in the Negative Face (**A**) and Neutral Face (**B**) groups for the Test in Original Context (blue), Test in Similar Context (red), and Test in New Context (grey) conditions. The scores are plotted separately for the Same Context Exposure (left), Similar Context Exposure (middle), and No Exposure (right) groups. Error bars represent SEM.