Memory transience versus memory persistence

Cox, W.R.

Publication date
2022

Citation for published version (APA):
Cox, W. R. (2022). *Memory transience versus memory persistence*.

CHAPTER 2

Time-dependent emotional memory transformation: divergent pathways of item memory and contextual dependency

ABSTRACT

Emotional memory can persist strikingly long, but it is believed that not all its elements are protected against the fading effects of time. So far, studies of emotional episodic memory have mostly investigated retention up to 24h post-encoding, and revealed that central emotional features (items) are usually strengthened, while contextual binding of the event is reduced. However, even though it is known for neutral memories that central versus contextual elements evolve differently with longer passage of time, the time-dependent evolution of emotional memories remains unclear. Hypothetically, compared to neutral memories, emotional item memory becomes increasingly stronger, accompanied by accelerated decay of – already fragile – links with their original encoding contexts, resulting in progressive reductions in contextual dependency. Here, we tested these predictions in a large-scale study. Participants encoded emotional and neutral episodes, and were assessed 30 minutes (N = 40), one day (N = 40), one week (N = 39), or two weeks (N = 39) later on item memory, contextual dependency, and subjective quality of memory. The results show that, with the passage of time, emotional memories were indeed characterized by increasingly stronger item memory and weaker contextual dependency. Interestingly, analyses of the subjective quality of memories revealed that stronger memory for emotional items with time was expressed in familiarity, whereas increasingly smaller contextual dependency for emotional episodes was reflected in recollection. Together, these findings uncover the time-dependent transformation of emotional episodic memories, thereby shedding light on the ways healthy and maladaptive human memories may develop.

1.   INTRODUCTION

Memories are not permanent records of past experiences, but can fade over the course of years, days, or even hours, after the original event. Not all memories of personally experienced events (i.e., episodic memories, Tulving, 1972) fade equally with the passage of time though. A wealth of empirical research emphasizes that especially emotionally arousing events are easily retrieved and recognized well after their initial occurrence (Hamann, 2001). However, despite their strong nature, emotional memories likely do not remain stable, preserved exactly in their original form. For example, it is evident that conditioned fear memories are highly specific one day after learning but gradually become more generalized (Wiltgen & Silva, 2007). Yet, it remains largely unknown how exactly emotional episodic memories may transform over time.

Most of what is understood about long-term changes of episodic memories stems from studies that have used emotionally neutral material (Cox et al., 2014; Sekeres et al., 2016; Talamini & Gorree, 2012). These studies typically show that the central feature of an event (often referred to as *item memory*) is remembered longer than associations between the item and the surrounding context (*contextual memory*). This preferential loss of contextual information has been attributed to the process of systems consolidation, in which the slow transfer of memory representations from the hippocampus to the neocortex is accompanied by memory schematization and decay of contextual memory (Barry & Maguire, 2019; Frankland & Bontempi, 2005; Tonegawa, Morrissey, & Kitamura, 2018; see Meeter & Murre, 2004 and Moscovitch, Cabeza, Winocur, & Nadel, 2016 for reviews of alternative theories). During this time-dependent process, contextual memory representations in the hippocampus are thought to be overwritten rather quickly (McClelland et al., 1995). In contrast, item memory is thought to be mediated by extra-hippocampal brain regions where memories decay at a slower rate (Talamini & Gorree, 2012).

As opposed to neutral episodic memory, so far, long-term fading of emotional items and their contexts remains poorly understood, although some studies did investigate these memory characteristics relatively shortly after learning. At such recent time points (i.e., immediately after learning, or 24h later), emotional episodes are characterized by strong item memory but weak contextual memory, in comparison to neutral episodes (Bisby & Burgess, 2014; Kensinger, 2009; Kensinger et al., 2007; Matsumoto & Kawaguchi, 2020; van Ast et al., 2013, 2014; Yonelinas & Ritchey, 2015). This co-occurrence of enhanced item memory and impaired contextual memory directly after learning is believed to be an important factor in the development of trauma intrusions and fear generalization: Strong emotional item memory that is just loosely bound to the encoding context may be readily cued in irrelevant contexts, such that disproportionate emotional responses can occur in safe situations (Brewin et al., 2010; Ehlers & Clark, 2000). Weak contextual dependency of recently encoded memories has indeed been found to be associated with more frequent analogue trauma intrusions (Bisby et al., 2010; Meyer et al., 2017) and more severe reexperiencing symptoms in posttraumatic

stress disorder (Sierk et al., 2019). However, emotional memories may remain subject to ongoing transformations much further in time. In rodents, fear generalization can accelerate over the course of several weeks after conditioning, which points to important changes of emotional memories that drive excessive fear responding well beyond initial learning (Biedenkapp & Rudy, 2007; Jasnow et al., 2017; Poulos et al., 2016; Wiltgen & Silva, 2007). Insight into time-dependent transformations of human emotional episodic memories could therefore have great relevance for understanding the alterations of healthy episodic memories, but may also shed light on the development of emotional memory generalization and intrusions.

Neurobiologically, strengthened memory for emotional items is thought to be mediated by upregulated amygdala activity during and after memory encoding (see Yonelinas & Ritchey, 2015 for a review). Memory for associations between separate items and between items and context is mediated by hippocampal-dependent activity (Burgess et al., 2002; Davachi, 2006), and undergoes an emotion-induced impairment related to downregulation of the hippocampus (Berkers, Klumpers, & Fernández, 2016; Bisby, Horner, Hørlyck, & Burgess, 2016; Murray & Kensinger, 2014; Okada et al., 2011, *but see* Madan, Fujiwara, Caplan, & Sommer, 2017). For item memory, the rapid emotion-induced enhancement by amygdalar upregulation results in not only high initial strength, but also a particularly slow decay. As time passes, emotional memories should therefore become increasingly characterized by strengthened item memory (Yonelinas & Ritchey, 2015). Conversely, contextual memory of emotional episodes might show accelerated time-dependent weakening. Namely, the already fragile contextual memory representations immediately after encoding might be particularly sensitive to overwriting processes in the hippocampus (McClelland et al., 1995). Also, since systems consolidation is believed to be a competitive process in which weak memories are sacrificed in favor of the endurance of stronger ones (Frankland & Bontempi, 2005; Meeter & Murre, 2004; Talamini & Gorree, 2012), contextual memory of emotional events might be transferred to the neocortex to a particularly small degree and therefore decay quickly. Hence, with time, emotional episodic memories may become increasingly characterized by strong item memory and weak contextual memory.

These conceivable long-term changes of emotional item and contextual memory likely become evident in specific subjective memory qualities with which events are remembered. Two separate subjective qualities of remembering have been characterized: a sense of recollection that includes vivid reexperiencing of the learning event, and a sense of familiarity where no specific associations come to mind (Yonelinas & Levy, 2002). Several studies suggest that enhanced recollection – not familiarity - underlies increasingly better memory for emotional than neutral items across short time frames (e.g., on the same day versus one day after learning, see Yonelinas & Ritchey, 2015 for a review). Furthermore, contextual dependency of episodic memory has sometimes been shown to manifest itself in altered recollective experiences (Macken, 2002). Therefore, both increasingly stronger retention of emotional items, as

well as accelerated loss of contextual dependency for emotional episodes, would likely be reflected specifically in recollection, but not familiarity.

Thus, in the present study, we aimed to test whether emotional episodes show, relative to neutral episodes, (i) retainment of item memory, and (ii) loss of contextual dependency, that both become increasingly pronounced with the passage of time. Furthermore, we assessed whether (iii) the stronger retainment of emotional items and (iv) the faster loss of contextual dependency for emotional episodes are specifically expressed in recollective experiences. We tested these predictions in a large-scale study where participants encoded emotional and neutral events and performed memory tests at one of several later time points. Specifically, during encoding, participants were shown neutral and negative words on white backgrounds and were instructed to focus on the meaning of the words, to eventually assess item memory. Participants were also presented with neutral and negative words on background pictures, for a later assessment of contextual dependency. They were instructed to come up with vivid scenes that include the words in the context images (van Ast et al., 2013). Time-dependent changes in item memory and contextual dependency of neutral and emotional episodes were determined by allocating participants to groups that performed memory testing 30 minutes, one day, one week, or two weeks after encoding. To assess item memory, words that were presented without a context image during encoding were again displayed on a white background for cued recall and recognition tests. To assess contextual dependency, half of the words that were initially presented with a context image were tested on the same background as during encoding, while the other half was shown on a different context image during testing (Cox et al., 2014; van Ast et al., 2013, 2014). In addition, we assessed the subjective qualities of memories by having participants perform a remember/know/guess procedure (Eldridge et al., 2002; Tulving, 1984).

2. MATERIALS AND METHODS

2.1 Participants. Dutch-speaking participants, aged between 18 and 40 years, participated in exchange for course credit or €20, -. The inclusion criteria were: no psychiatric or neurological disorder requiring psychological or pharmacological treatment and no diagnosis of dyslexia. Participants were excluded if they had previously participated in similar experiments. The local ethical committee of the University of Amsterdam approved all procedures.

A total of 181 participants consented to take part. Twenty-three participants met one of the exclusion criteria (psychiatric disorder, n = 13; prior participation in similar research, n = 5; dyslexia, n = 1). Furthermore, three participants encountered a technical error, and one prematurely ended his participation. Consequently, we analyzed data of 158 participants (120 females) with a mean age of 22 years (SD = 3.53, range = 18 - 38). The participants were allocated to one of four groups that performed item and contextual dependency tests either on the same day as encoding (n = 40), one day later (n = 40), one week later (n = 39), or two weeks later (n = 39).

2.2   Stimuli

2.2.1.   Words. The words for the memory task were 120 nouns obtained from a validated database of Dutch words (Moors et al., 2013), as has been used in previous studies (van Ast et al., 2013, 2014). The words consisted of three to ten letters, and contained no more than four syllables. Half of the words were of neutral valence (range 3.09 − 4.69, total scale = 1 − 7), and the other half of negative valence (range = 1.16 − 2.77, total scale = 1 − 7). For the creation of different task versions, the words were assigned to 12 sets of 10 words each (i.e., 6 neutral word sets and 6 negative word sets). The average valence scores of all individual negative word sets were significantly lower than those of all individual neutral word sets, as assessed by a univariate ANOVA ($F_{11,108}$ = 81.42, p < .001) and Tukey HSD tests (all p < .001). The negative word sets did not differ in valence from each other ($F_{5,54}$ = 1.39, p = .242). Likewise, there was no significant difference in valence between the neutral word sets ($F_{5,54}$ = 1.07, p = .387). In addition, the 12 sets were matched such that they did not differ in word length ($F_{11,108}$ = 0.59, p = .837), and familiarity ($F_{11,108}$ = 1.23, p = .280).

2.2.2.   Contexts. Context images were in total forty color pictures of indoor and outdoor spatial environments (e.g., a living room, city landscape, van Ast et al., 2013, 2014).

2.3   Experimental task

2.3.1.   Encoding. The experimental paradigm (Fig. 1) was based on Cox et al. (2014) and van Ast et al. (2013, 2014). During encoding, in total 60 words were presented on a computer screen, of which 10 neutral (1 set) and 10 negative (1 set) were displayed on a white background, and 20 neutral (2 sets) and 20 negative (2 sets) words in a background picture. The 40 words in a context image were paired with the background pictures in such a way that each word set contained an equal number of indoor and outdoor spatial environments. Each trial started with the background (white or picture) being displayed for 1 second, after which the word was shown in the middle of the background for 5 seconds (see 2.3.4 Trail order for more information). Then the word disappeared, and the background was shown alone for 1 additional second. Subsequently, participants rated the arousal and valence of the imagined scene, using self-assessment manikins (Bradley & Lang, 1994). For each rating, participants were given four seconds to respond during which their reaction times were recorded. The next trial started after presentation of a black background screen for 1 second (intertrial interval). The participants were presented with three practice trials before the start of the task.
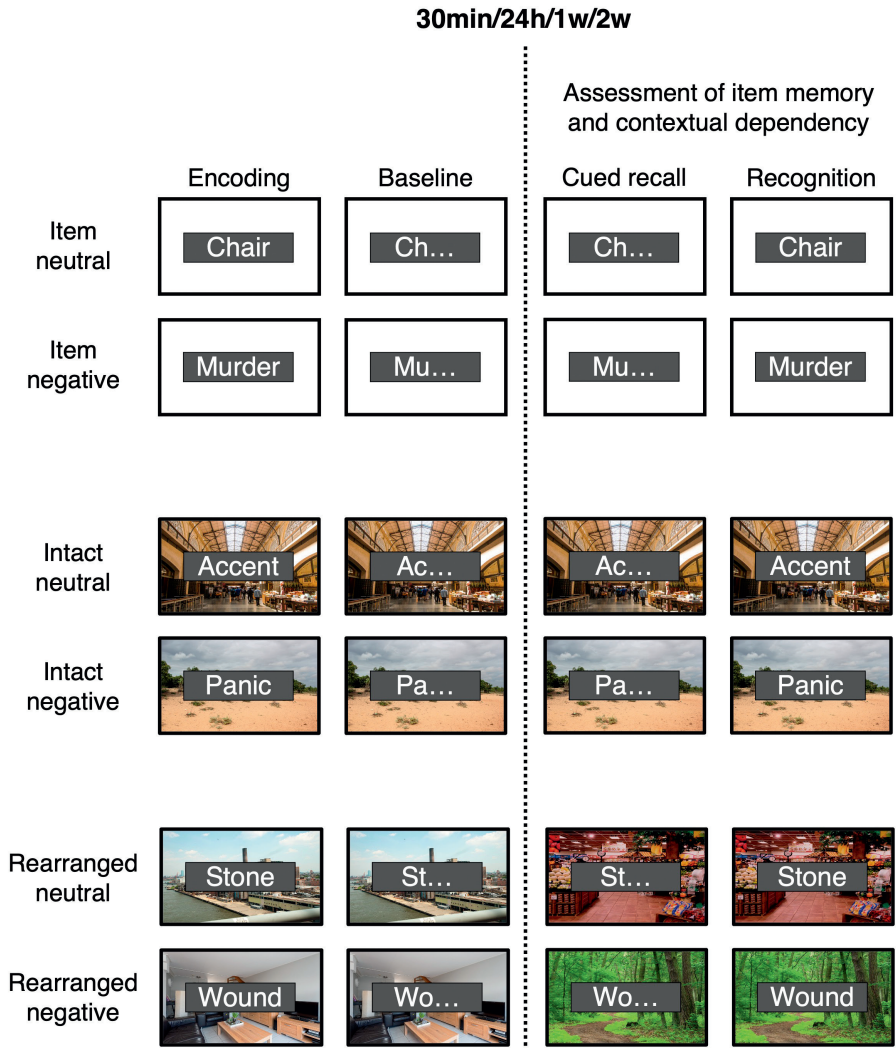
Fig. 1. Experimental paradigm. During encoding on day 1, neutral and negative words were presented on background pictures or white backgrounds, and presented intermixed. Participants vividly imagined scenes with the words in the context images, or just the words in case of a white background. Afterwards, participants performed a baseline cued recall test. Here, the first two letters were presented as cues on the same background image as during encoding. Thirty minutes, one day, one week, or two weeks later, participants returned and completed another cued recall test, followed by a recognition test. Crucially, the cues (for cued recall) or complete words (for recognition) were now presented either in the same context as during encoding (intact), or in a different context (rearranged). For the words that had been encoded without contexts, a white background was again presented. During recognition an equal number of new words was shown as well (not displayed here).

2.3.2. Baseline cued recall test. After encoding of all 60 words, participants completed a self-paced cued recall test. Specifically, the first 2 letters of each word were presented on the same background (white or picture) as during encoding to cue the memory. Each trial started with the background being displayed for 1 second, after which the word cue appeared and stayed on the screen until participants had typed in their response and pressed enter. Reaction times were recorded. The next trial started after the presentation of a black screen for 1 second.

2.3.3. Assessment of item memory and contextual dependency.
2.3.3.1. Cued recall. During the second session, participants again performed a self-paced cued recall test. For the assessment of item memory, the cues of the 20 words that were encoded without a context image were - like during baseline cued recall - presented on a white background (item$_{neutral}$, item$_{negative}$). The 40 words that had been presented in a picture during encoding were used to assess contextual dependency. Crucially, the letters of the words appeared in the same context image as before for half of the trials (1 neutral set, 1 negative set), whereas for the other half (1 neutral set, 1 negative set) the combination of context images and cues were randomly shuffled, such that the letters appeared in a different - yet earlier seen – context. In doing so, we created two conditions in which contextual information was present to support memory recall (intact$_{neutral}$, intact$_{negative}$), and two conditions that also included contextual information but not for the presented word (rearranged$_{neutral}$, rearranged$_{negative}$). The timing of individual trials was the same as during the baseline cued recall test.

2.3.3.2. Recognition. During the subsequent recognition test, all 60 words that participants had seen before, and 60 new words (lures) were presented. Specifically, to assess item memory, the 20 old words that had been encoded without a context image were again presented on a white background. Also, 20 new words (1 neutral set, 1 negative set) were presented without a context image. For the assessment of contextual dependency, the 40 old words that had been paired with a context – like for the previous recall test - were either shown on the same background image as during encoding, or on a different image. The combination of contexts and words in the rearranged conditions were shuffled anew, so that they were not the same as in the cued recall test. Furthermore, 40 new words (2 neutral sets, 2 negative sets) were presented on one of the context images that also had been presented with one of the old words. The new words were assigned to a background image in such a way that the contexts that were paired with the intact$_{neutral}$, intact$_{negative}$, rearranged$_{neutral}$, and rearranged$_{negative}$ within-subject conditions were equally distributed over the new neutral, and new negative words.

Each trial started with the background being displayed for 1 second, after which the word appeared and stayed on the screen until participants had indicated whether they recognized the word (old) or not (new). To assess the subjective qualities of recognition (i.e., recollection and familiarity), participants were, if they responded "old", shown three response option (remember, know guess) of which they could select one

(Eldridge et al., 2002). If participants responded "new", they were instead presented with three options to indicate their confidence in the response (sure, not sure, guess). We included this confidence rating task to ensure that the number of responses would be equal for both recognition responses (old/new), so that responding "new" would not enable participants to shorten the duration of the experiment (van Ast et al., 2013). These confidence ratings were thus used as a filler task only and not further analyzed. Reaction times of old/new, remember/know/guess, and sure/not sure/guess responses were recorded. The next trial started after presentation of a white fixation cross on a black background image for 1 second.

2.3.4. Trial order. The order of trial presentation during all phases of the experiment was semi-randomized with the restriction that words from the same within-subject condition (e.g., $item_{neutral}$, $item_{negative}$, $intact_{neutral}$, $intact_{negative}$, $rearranged_{neutral}$, $rearranged_{negative}$) would not be presented more than two times consecutively. Furthermore, the trials that were presented during the first half of the encoding task (i.e., the first 30 trials) were also presented in the first half of the memory tests. We included this separation of trials in two blocks as we anticipated potential floor effects in the cued recall test that was performed during the second session. That is, if memory performance in one or several groups would be low overall (most likely in the one week and two weeks conditions), we could validly analyze the first half of the trials that might show relatively improved retention due to primacy effects.

2.3.5. Experimental task versions. We created 6 versions of the experimental task that were equally divided across the groups. Across all versions, the same 60 words (3 neutral sets, 3 negative sets) were presented during encoding and all tests, whereas the other 60 words (3 neutral sets, 3 negative sets) were always used as lures during the recognition test. To rule out that any effects are driven by specific word sets, we made sure that the word sets were assigned to different-within subject conditions across the versions. Furthermore, in each version, the context images were assigned to the words in the intact and rearranged context conditions in a random fashion.

2.4 Procedure. Upon arrival on day 1, as a cover story, participants were told that the purpose of the experiment was to study practice effects of an imagination task. They were informed that the imagination task involved words - including negatively valenced ones - and background images. It was stated that they would perform an imagination task during both sessions. We deliberately kept participants unaware that they would perform memory tests to preclude active memorizing of the imagined scenes in between sessions. Subsequently, participants gave written informed consent and were screened for the exclusion criteria. Before the start of the experimental task, participants filled in questionnaires. These data were collected for potential exploratory analyses, but not analyzed.

2.4.1. Encoding task. Participants were seated in front of a computer in a dimly lit room, and were explained that they would be shown words depicted against background images. They were instructed to come up with a vivid scene that involves the meaning of the word in the context image while the word and the image were on the screen (van Ast et al., 2013). They were also informed that some of the words would appear on a white background, instead of a context image. Participants were asked to imagine the content and meaning of these words without the background. It was explained that after presentation of the word and the image, they would rate the valence and arousal of their imagined scenes.

2.4.2. Baseline cued recall test. Upon completion of the encoding task, participants were explained that the first two letters of the words they had just seen would be presented as cues on the same background images as during encoding. They were instructed to finish the words by typing in the missing letters. If they did not remember the correct word, they were asked to type in "xx". Upon completion of the baseline cued recall test, participants were informed that they would perform a similar imagination task in the second session. Participants in the same day condition were given a short break (i.e., half an hour), before starting the second session.

2.4.3. Assessment of item memory and contextual dependency.
2.4.3.1. Cued recall. Upon their return to the laboratory, participants were explained that they would perform a similar cued recall test as during the first session. They were informed that during this test, however, the background image presented with the cue letters would not always be the same as during encoding, but could sometimes also be a different one. The other instructions were the same as during the baseline cued recall test.

2.4.3.2. Recognition. Next, participants were explained that they would be presented with complete words they had seen during the first session and words they had not seen previously. They were asked to indicate whether they recognized the word as one that had been presented before (old) or not (new). It was made clear that if they recognized the word, they would subsequently be asked to indicate if they actively remembered the word or whether their response was based on a sense of knowing instead. The difference between remembering and knowing was explained using validated instructions (Geraci et al., 2009; Rajaram, 1993). In summary, participants were asked to give a "remember" response if the word elicited a conscious recollection of specific thoughts or images from when they encoded the word, and a "know" response if they were certain that they had seen the word before but no specific associations came to mind. They were informed that they also had the option of indicating that their response was a guess (Eldridge et al., 2002). In addition, participants were explained that if they did not recognize the word (new), they would then indicate the level of confidence in their response (van Ast et al., 2013)

Upon completion of the recognition test, participants were interviewed regarding encoding characteristics, such as the use of background picture while imagining the scenes. These data were collected for potential exploration but not further analyzed. Finally, participants were debriefed and rewarded for participation.

2.5   Data analysis. The Statistical Package for the Social Sciences (SPSS) version 25.0 (Armonk, NY: IBM Corp.) was used to preprocess and analyze the data. The preprocessing and analyses of valence and arousal ratings, as well as the cued recall data (baseline and assessment of time-dependent effects) are described in Supplementary Methods. The data and analysis code have been deposited at Open Science Framework (OSF) and are available at https://osf.io/emxgv/. The study was not preregistered.

2.5.1.   Preprocessing for assessment of item memory and contextual dependency.
2.5.1.1. Recognition. Hit rates were calculated for each of the within-subject conditions ($item_{neutral}$, $item_{negative}$, $intact_{neutral}$, $intact_{negative}$, $rearranged_{neutral}$, $rearranged_{negative}$), and false alarm rates for neutral and negative lure words per background type (picture, white). From these hit and false alarm rates, we computed d-prime sensitivity index. For this purpose, hit rates of 1 were truncated at 0.95, and false alarms rates of 0 at 0.025 (picture) or 0.05 (white)[2]. Next, contextual dependency scores for neutral and negative memories were calculated in two ways. First, we calculated contextual dependency using d-prime difference scores: $context_{recognition} = intact_{recognition} - rearranged_{recognition}$. We also calculated contextual dependency hit rate scores based on proportions: $context_{hits} = 1 - (rearranged_{hits} / intact_{hits})$. This second measure was included because initial differences between the intact and the rearranged context conditions could confound interpretations of time-dependent changes of contextual dependency when difference scores are used. That is, a higher accuracy score in the intact versus rearranged context condition in the same day group could lead to a differential decrease in accuracy over time (i.e., a larger decrease in the intact than the rearranged condition due to baseline differences), such that concluding a loss of contextual dependency would be invalid. As the false alarm rates are the same for the intact and rearranged context conditions (presentations of new words are not related to context manipulations from encoding to testing), and when false alarm rates are constant d-prime is a nonlinear measure of memory accuracy (Stanislaw & Todorov, 1999), proportional d-prime scores to assess contextual dependency could not be calculated to circumvent this issue. However, we could use hit rates for this purpose as this *is* a linear measure of memory accuracy (proportional scores can thus be calculated to control for initial

---

2   To avoid invalid scores, it is recommended to replace values of 1 with (n − 0.5)/n and values of 0 with 0.5/n, where n refers to the number of trials (Stanislaw & Todorov, 1999). As hit rates of all within-subject conditions are based on 10 trials, 1 is replaced with 0.95. As false alarms of items in context images are based on 20 trials, but in white backgrounds on 10 trials, 0 is replaced with 0.025 and 0.05, respectively.

differences between the context conditions), and these scores are based on the same data as d-prime (since false alarms rates are the same for the context conditions, both measures only include hit rates). As a final step to obtain the variables for the analyses, we computed difference scores between negative and neutral memories for both item and contextual dependency (neutral subtracted from negative).

2.5.1.2. Recollection and familiarity. Recollection and familiarity scores were calculated for all within-subject conditions ($item_{neutral}$, $item_{negative}$, $intact_{neutral}$, $intact_{negative}$, $rearranged_{neutral}$, $rearranged_{negative}$) using the formulas provided by (Yonelinas et al., 1998). For recollection, this is: ($remember_{hits}$ - $remember_{false\ alarms}$) / (1 - $remember_{false\ alarms}$). Familiarity scores were calculated from an adjusted signal detection procedure: d' = $know_{hits}$ / (1 - $remember_{hits}$) versus $know_{false\ alarms}$ / (1 - $remember_{false\ alarms}$). Contextual dependency of memory recollection was calculated using proportional scores: $context_{recollection}$ = 1 - ($rearranged_{recollection}$ / $intact_{recollection}$). For contextual dependency of familiarity-based recognition, we calculated difference scores, and hit rate proportional scores, similar to the d-prime recognition contextual dependency scores: $context_{familiarity}$ = $intact_{familiarity}$ - $rearranged_{familiarity}$, and $context_{familiarity}$ = 1 - ($rearranged_{knowhits}$ / (1 - $rearranged_{rememberhits}$)) / ($intact_{knowhits}$ / (1 - $intact_{rememberhits}$)). Again, differential item memory and contextual dependency scores (neutral subtracted from negative) were calculated for both recollection and familiarity.

2.5.2. Statistical analyses of item memory and contextual dependency.
2.5.2.1 Recognition. For our primary analyses, ANOVA linear contrast tests were performed, in which Time (same day, one day, one week, two weeks) was entered as independent variable. To assess the predictions of better retainment of emotional item memory and stronger decay of emotional contextual dependency that both become more pronounced with time in comparison to neutral memories, we performed these linear contrast analyses using the recognition difference scores (negative -neutral, see 3.2 primary results). We used -3, -1, 1, and 3 (item), or 3, 1, -1, -3 (context) as contrast weights, reflecting a positive relationship and a negative relationship with the passage of time for the item memory and contextual dependency difference scores, respectively. Note that such a contrast analysis is well suited to test one specific hypothesized pattern across time, whereas a significant effect in a conventional ANOVA may reflect a (unpredicted) difference between just two time points (Haans, 2018).

As secondary analyses, we investigated if general reductions over time occurred in item memory and contextual dependency (Cox et al., 2014; Talamini & Gorree, 2012), and if overall stronger item memory and weaker contextual dependency could be observed for emotional versus neutral memories (Yonelinas & Ritchey, 2015). Specifically, we tested for effects of Time (using the same contrasts weights as described above) and a main effect of Emotion on the item (negative and neutral) and context (negative and neutral) scores (see 3.3 secondary results).

2.5.2.2. Recollection, and familiarity. We performed these same (primary) analyses with the recollection and familiarity difference scores to test if stronger retainment of emotional items than neutral items, and a faster loss of contextual dependency for emotional versus neutral episodes, is specifically expressed in memory recollection.

3.   RESULTS

3.1.   Manipulation checks. Analyses of valence and arousal scores confirmed that the creation of negative versus neutral episodes had succeeded, and revealed that no significant differences in subjective ratings of valence and arousal occurred between the groups. Furthermore, effects of negative versus neutral words on valence and arousal were found to depend on background (picture, white) (see Supplementary Results, and Fig. S1 for more details).

The baseline cued recall data confirmed that there were no significant differences in baseline accuracy between the groups (see Supplementary Results). Additionally, we observed better memory for negative items than neutral items, and for items in context images as compared to items in white backgrounds (Fig. S2).

3.2.   Primary results.
3.2.1.   Cued recall. No time-dependent differences were found between negative and neutral memories in either item memory or contextual dependency (see Supplementary Results).

3.2.2.   Recognition. In line with the predictions, the contrast analysis of d-prime item memory difference scores (item$_{negative}$ - item$_{neutral}$, white backgrounds) confirmed that negative items were recognized increasingly more accurate over time, relative to neutral items (C = 2.32, 95% CI [1.22, 3.43], $F_{1,154}$ = 17.31, p < .001, η_p^2 = .10, Fig. 2A). Conversely, contextual dependency of negative episodes compared to neutral episodes (Context$_{negative}$ - Context$_{neutral}$) indeed became increasingly smaller over time (C = -1.17, 95% CI [-2.06, -0.27], $F_{1,154}$ = 6.61, p = .011, η_p^2 = .04, Fig. 2B, S6).

Analyses of hit rates mirrored these effects (Supplementary Results, Fig. S4, S7), ruling out that the contextual dependency results can be explained by differences between the intact and rearranged context conditions at the first time point (i.e., the same day condition). Note that generally high hit rates in the same day and one day conditions could have contributed to the gradual development of higher contextual dependency scores over time (see 3.3 secondary results).
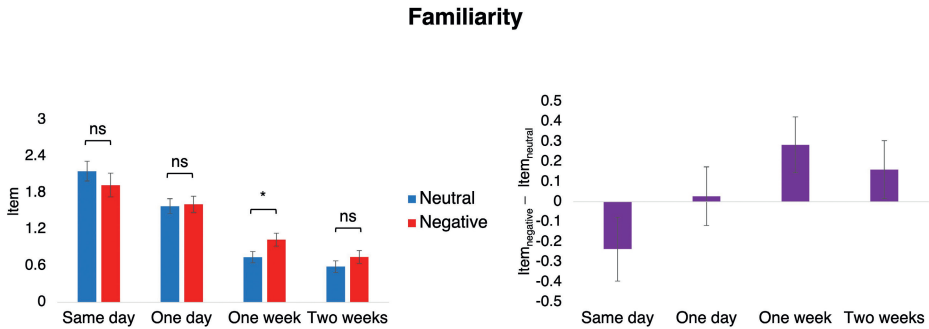
D-prime accuracy
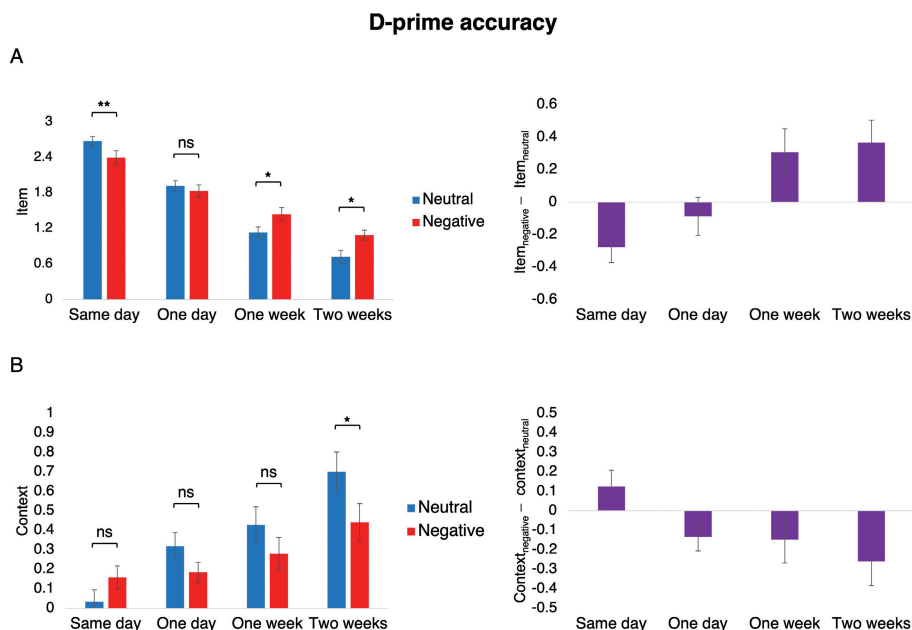
Fig. 2. (**A**) Item memory results displayed for neutral and negative memories separately (left), and differences scores (negative – neutral, right). We observed a stronger retainment of negative items versus neutral items that became more pronounced with time. (**B**) Contextual dependency results displayed for neutral and negative memories separately (left), and differences scores (negative – neutral, right). Over time, contextual dependency of negative memories became increasingly smaller relative to neutral memories. * $p < .05$, ** $p < .01$, ns = not significant. Error bars represent SEM.

3.2.3. Recollection and familiarity. The analyses of subjective memory ratings did not reveal convincing evidence for the hypothesis that better retention of negative items versus neutral items over time is expressed in the subjective experience of recollection ($C = 0.28$, 95% CI [-0.03, 0.60], $F_{1,154} = 3.10$, $p = .080$, $\eta_p^2 = .02$, Fig. 3A). Contrary to our predictions, the time-dependent enhancement of negative versus neutral items was in fact significantly expressed in measures of familiarity ($C = 1.45$, 95% CI [0.12, 2.77], $F_{1,147} = 4.67$, $p = .032$, $\eta_p^2 = .03$, Fig. 4).

Furthermore, in line with our expectations, we observed that the loss of contextual dependency for negative episodes compared to neutral episodes was expressed in recollection ($C = -1.06$, 95% CI [-1.94, -0.19], $F_{1,150} = 5.76$, $p = .018$, $\eta_p^2 = .04$, Fig. 3B, S7). In this analysis, four participants (two in the same day condition, one in the one week condition, and one in the two weeks condition) were excluded due to invalid scores (i.e., minimum recollection in the intact context condition leading to dividing by zero). The assessment of contextual dependency for familiarity-based recognition was compromised by missing data in a large proportion of participants (0.34), most notably in the same day (0.58), and one day conditions (0.43). This was due to remember$_{hits}$ scores of 1, such that familiarity scores were zero. Therefore, the familiarity contextual dependency scores were not analyzed.
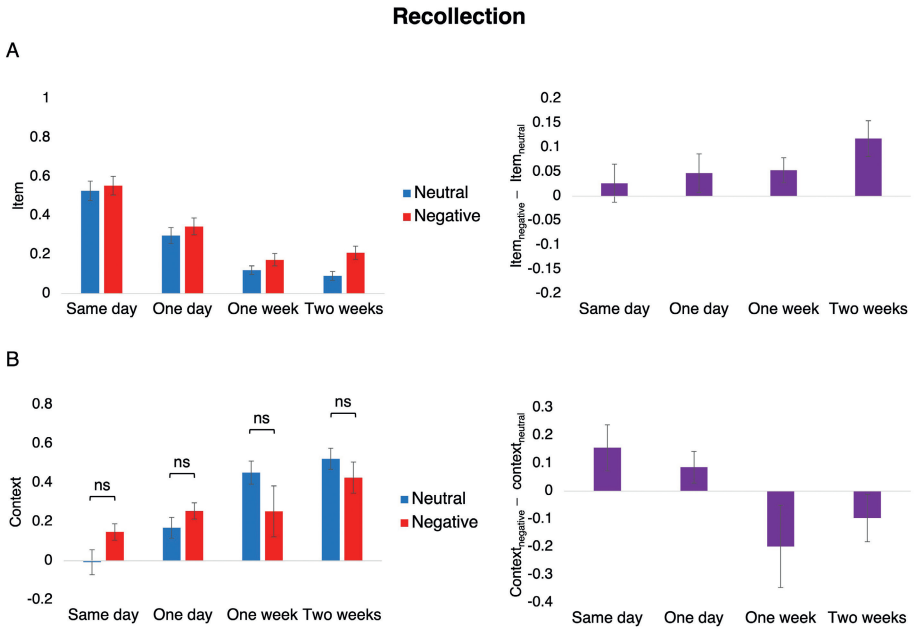
Recollection

Fig 3. (**A**) Results of item recollection displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). No significant retainment of negative versus neutral items that becomes increasingly pronounced over time was observed. (**B**) Assessment and results of contextual dependency of recollection displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). Over time, contextual dependency of negative memories became increasingly smaller compared to neutral memories. ∗ *p* < .05, ns = not significant. Error bars represent SEM.

Familiarity

Fig. 4. Results of item familiarity displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). A significant retainment of negative versus neutral items that becomes increasingly pronounced over time was observed. ∗ *p* < .05, ns = not significant. Error bars represent SEM.

3.3.   Secondary results. Regardless of emotion, item memory d-prime scores generally decreased with time ($F_{1,154}$ = 261.35, p < .001, η_p^2 = .63, Fig. 2A). Surprisingly, contextual dependency, conversely, showed a time-dependent gradual enhancement ($F_{1,154}$ = 31.89, p < .001, η_p^2 = .17, Fig. 2B, S6). The hit rate data show that this development of larger contextual dependency scores over time could be due to near-maximal hit rates in the same day group and the one day group. That is, contextual dependency scores were particularly small in the groups that performed the memory tests relatively shortly after learning (same day and one day groups) as performance was high in both the intact and rearranged context conditions, whereas for the later groups (one week and two weeks) overall performance was lower such that better memory in the intact versus rearranged context condition emerged (Fig. S8). False alarm rates were not minimal, however, such that recognition was not perfect in any group, and thus memory itself was not at ceiling (Fig. S5).

Furthermore, d-prime accuracy was overall not higher for negative items than neutral items ($F_{1,154}$ = 1.54, p = .217, η_p^2 = .01). This is likely due to the same day condition, interestingly, showing superior recognition for *neutral items*, rather than negative items (t_{39} = -2.91, p = .006), whereas at later time points the pattern reverses such that *negative items* are recognized better (one day: t_{39} = -0.75, p = .461, one week: t_{38} = 2.12, p = .041, two weeks: t_{38} = 2.66, p = .011, Fig. 2A). Finally, contextual dependency of negative memories was overall indeed lower than neutral memories ($F_{1,154}$ = 4.19, p = .042, η_p^2 = .03, Fig. 2B, S6).

4.   DISCUSSION

Overall, the findings confirm that time-dependent development of item memory and contextual dependency is notably different for negative memories compared to neutral memories. With the passage of time, recognition of negative items versus neutral items indeed becomes increasingly accurate. The analyses of memory in context, conversely, suggest a relative reduction in contextual dependency for negative episodes when time passes. In contrast with our hypothesis, the increasingly stronger negative item memory with time was expressed in increases in familiarity, whereas this same analysis for recollection was only marginally significant. We also found that the reduced contextual dependency for negative episodes over time was expressed in recollection, in line with our predictions. Together, these findings shed light on the way episodic memories develop on the long term and how they may guide behaviour after the occurrence of an emotional event.

The observation that negative items are retained to a larger degree than neutral items is consistent with many earlier studies showing generally enhanced memory for emotional materials at shorter time intervals (Hamann, 2001). Data addressing a *time-dependent* benefit for emotional items in human episodic memory research is, however, surprisingly scarce. Previous studies in which emotional and neutral item memory were found to differ after a delay are often limited to tests performed immediately after learning and one day later (Sharot & Phelps, 2004; Sharot & Yonelinas, 2008).

Furthermore, studies that included longer learning-test intervals often did not find time-dependent differences in accuracy between emotional and neutral items (e.g., Ritchey et al., 2008; Wang, 2014, 2018; Wang & Fu, 2011; Weymar et al., 2011, but see e.g., Anderson et al., 2006). This could be due to relatively small samples (Ritchey et al., 2008; Weymar et al., 2011), or repeated testing of the same learned material (Weymar et al., 2011), which can bias results (Burke et al., 1992). Also, in some studies interaction ANOVA tests were performed (Wang, 2014, 2018; Wang & Fu, 2011). Such an analytic strategy is not well suited to assess time-dependent changes, because of its sensitivity to differences in accuracy at *any* time point instead of a specifically predicted gradual change with time (for which contrast analysis is needed, Haans, 2018). The large sample size in the present study (N = 158), and the between-subjects design in which separate groups performed tests at different time points, likely contribute to the reliability of the results, showing that retainment of emotional items does grow stronger with time. This finding cannot be explained by simply greater attention to emotional than neutral stimuli, because on the same day as learning recognition of negative items was lower – not higher – than neutral items. It was only with time that negative items developed a memory advantage compared to neutral items, suggesting that post-encoding memory processes underlie this effect. This observation is in line with traditional consolidation theory (McGaugh, 2000), as well as contemporary alternative accounts of memory storage, such as emotional binding theory (Yonelinas & Ritchey, 2015), which both predict increasingly stronger memory for emotional items with the passage of time.

With regard to contextual memory, we presumed that the transfer of memory representations from the hippocampus to the neocortex (systems consolidation) would occur to a relatively small degree for emotional episodes, such that these memories are characterized by increased reductions in contextual dependency with time. Emotional versus neutral episodes indeed displayed increasingly smaller contextual dependency with the passage of time. However, we did not observe a general time-dependent decrease of contextual dependency. We instead observed a general increase, which seems to indicate that a different process than systems consolidation is at play. How can these observations be explained? The data show that recognition was initially (i.e., in the same day and one day conditions) strongly item-based for both emotional and neutral memories. Since item memory was strong, context could therefore aid little in recognition, such that contextual dependency was low. Only after item memory had somewhat decreased (i.e., in the one week and two weeks conditions) there was room for the memories to become dependent on their encoding context. It seems that only at this point – when item memory was not dominating anymore – smaller contextual dependency for emotional versus neutral memories could be expressed. For this reason, proposed mechanisms for low contextual dependency of recent emotional memories (e.g., emotional arousal impairing contextual memory processing in the hippocampus, Bisby et al., 2016, or neutral episodes profiting from extra-hippocampal associative encoding, Madan et al., 2017) are plausible explanations for our findings. Attentional processes during encoding (Easterbrook, 1959; Kensinger et al., 2007; Loftus et al.,

1987) are, however, not a straightforward explanation as these often result in a memory trade-off immediately after learning (improved item memory accompanied by impaired contextual memory), which we did not observe. Note that some earlier studies have also assessed contextual memory of emotional versus neutral episodes at multiple time points. These studies mostly addressed incidentally encoded source information though (e.g., task instructions, Sharot & Yonelinas, 2008), which are usually not affected by the presence of emotional items (Yonelinas & Ritchey, 2015), and often did not include assessments beyond one day after learning (Waring & Kensinger, 2009). Our data reveal that in case of strong item memories, dependency on local context is initially low, and only likely to increase with the passage of time. Critically, this enhanced contextual dependency is slowed down for emotional memories. In subsequent studies it could be worthwhile to include even more remote time points and test if contextual dependency increases at first like we observe here, but eventually could decrease again (in line with systems consolidation theory), perhaps particularly for emotional memories (Barry & Maguire, 2019; Frankland & Bontempi, 2005; Tonegawa et al., 2018).

Interestingly, the observed time-dependent pattern of contextual dependency fits well with previously identified predictors of trauma intrusions and the way intrusions naturally develop over time. Our data show that shortly after learning, contextual dependency of well-learned items is low. Likewise, immediately after the experience of trauma, initial recognition of central trauma aspects conceivably does not require presence of the original encoding context either, and can be remembered in any potential context, given the strength of trauma memories. Since reduced contextual dependency of memory has been related to increased number and distress of analogue traumatic intrusions (Meyer et al., 2017; Voorendonk et al., 2021), the risk of distressing intrusions immediately after the occurrence of a traumatic event should therefore be high. Indeed, trauma analogue studies (Rattel et al., 2019) and clinical observations (O'Donnell et al., 2007) show that shortly after an emotional event reexperiences occur frequently. The present data show that over time however, contextual dependency, even of strong memories, clearly develops. Based on the relationship between contextual dependency and analogue intrusions (Meyer et al., 2017), the number of intrusions should hence decrease with the passage of time. Our study also reveals that emotional memories remain poorly contextualized relative to neutral memories at remote time points though, and may thus keep intruding to some extent even after a substantial delay. This is again in line with the analogue (Rattel et al., 2019) and clinical (O'Donnell et al., 2007) studies, revealing that with time intrusions may fade but not completely, thus following the pattern that one would predict based on the present contextual dependency data. For these reasons, it may be worthwhile to assess in future studies if time-dependent changes in contextual dependency of emotional memories indeed correlate with, or can even predict, the development of intrusions over time.

Not all aspects of memory were similarly affected by our manipulations. For example, the cued recall data did not show time-dependent differences in item memory or contextual dependency for emotional versus neutral episodes. Also, and crucial in light

of our hypotheses, the analyses of the subjective quality of recognition suggested that recollection and familiarity were not always impacted in the exact same way. Contrary to our hypothesis and earlier studies (Yonelinas & Ritchey, 2015), we found that the time-dependent enhancement of negative versus neutral items was expressed in stronger familiarity-based recognition, rather than in recollection (only a trend was observed in the predicted direction for recollection-based recognition). This discrepancy between the present work and earlier studies may be explained by differences in the presentation of emotional items: we displayed negative words on white backgrounds, whereas many previous studies presented emotional images during encoding (Anderson et al., 2006; Atienza & Cantero, 2008; Dolcos et al., 2005; Ochsner, 2000; Ritchey et al., 2008, 2019; Sharot et al., 2004; Wang, 2018). Emotional images contain negative objects within a specific environment, and thus in all likelihood contain more contextual information than words on a white background. Since retrieval of contextual memory is an inherent part of recollective experiences (Perfect et al., 1996), it is possible that strengthened memory for negative words that are presented on white backgrounds is particularly expressed in familiarity-based recognition due to the absence of contextual information. On the other hand, current theory (Yonelinas & Ritchey, 2015) and previous studies (Ritchey et al., 2019; Sharot et al., 2004) suggest that enhanced recollection of emotional versus neutral material is not so much driven by retrieval of contextual memory, but more so by enhanced item-emotion associations through involvement of the amygdala. As negative items were rated with higher arousal and lower valence than neutral items in our study, one would still expect a particularly strong effect on recollection and not necessarily familiarity. Hence, our data indicate that the role of emotion in familiarity-based item recognition may be worth investigating further to refine current theory, which assumes that enhanced memory for emotional items is by and large exclusively expressed in recollection (Yonelinas & Ritchey, 2015).

Our findings also show that time-dependent reductions in contextual dependency for emotional versus neutral memories is expressed in recollective experiences. This finding does align with previous research showing that effects of context on recognition is only observed when it is accompanied by conscious recollection (Macken, 2002). Note though that whether recollection or familiarity is impacted by contextual stimuli seems to depend on subtle conditions (e.g., response time, McKenzie & Tiberghien, 2004, and encoding instructions, Hockley, 2008). For this reason, and since contextual dependency of familiarity-based recognition could not be determined in the present study, it remains possible that familiarity is likewise or particularly impacted under different circumstances.

In conclusion, with this study we provide evidence that item memory and contextual dependency of emotional episodic memories, as well as their subjective qualities, show a markedly different development over time than neutral memories. Indeed, it seems that remote emotional memories are especially characterized by strong item memory and small contextual dependency. These findings shed important light on how our everyday memories of times past transform after learning. Furthermore, they point to processes that may lead to emotional memories intruding when in safe contexts.

SUPPLEMENTARY MATERIAL

1.   Supplementary Methods

1.1.   Data analysis.

1.1.1.   Preprocessing.

1.1.1.1. Manipulation checks. The average valence and arousal ratings of the participants' imagined scenes were calculated for each of the Background (picture, white) and Emotion (neutral, negative) within-subject conditions.

For baseline cued recall, all individual responses were inspected. Misspellings, singularity versus plurality mistakes, and diminutives of the target word were treated as correct responses. Like the valence and arousal ratings, the proportion of correct responses during the baseline cued recall test were calculated for the Background and Emotion within-subject conditions.

1.1.1.2. Assessment of item memory and contextual dependency: Cued recall. We first calculated the proportion of correct responses for all six within-subject conditions ($item_{neutral}$, $item_{negative,}$ $intact_{neutral}$, $intact_{negative}$, $rearranged_{neutral}$, $rearranged_{negative}$) per time point (same day, one day, one week, two weeks). Next, contextual dependency scores for neutral and negative memories were computed using the following formula: $context_{recall} = 1 - (rearranged_{recall} / intact_{recall})$. Finally, we calculated difference scores between negative and neutral memories for both item memory ($item_{negative} - item_{neutral}$), and contextual dependency ($context_{negative} - context_{neutral}$).

1.1.2.   Statistical analyses.

1.1.2.1. Manipulation checks. For both the valence and arousal ratings, as well the baseline cued recall data, a mixed ANOVA was performed with Background (picture, white) and Emotion (neutral, negative) as within-subject factors and Time (same day, one day, one week, two weeks) as between-subject factor. We performed planned contrasts to assess if negative items indeed resulted in visualizations with lower valence and higher arousal than neutral items.

1.1.2.2. Assessment of item memory and contextual dependency: Cued recall. The same analyses as for the recognition data were used (i.e., linear contrast analyses, see Main Manuscript).

2. Supplementary Results

2.1. Manipulation checks. Negative items elicited lower valence and higher arousal ratings than neutral items (Emotion, valence: $F_{1,152} = 1094.57$, $p < .001$, $\eta_p^2 = .88$, arousal: $F_{1,152} = 432.36$, $p < .001$, $\eta_p^2 = .74$, Fig. S1). These effects did not differ between the groups (Time × Emotion, valence: $F_{3,152} = 0.52$, $p < .670$, $\eta_p^2 = .01$, arousal: $F_{3,152} = 1.29$, $p = .279$, $\eta_p^2 = .03$). We also found that the differences between negative and neutral items in both valence and arousal scores were larger for items on context images than white backgrounds (Background × Emotion, valence: $F_{1,152} = 19.67$, $p < .001$, $\eta_p^2 = .12$, arousal: $F_{1,152} = 19.39$, $p < .001$, $\eta_p^2 = .11$). Nevertheless, for both background types the valence scores were lower - and arousal scores higher - for negative versus neutral items (all $t > 18.78$, all $p < .001$). Furthermore, any difference between items in context images versus items in white backgrounds did not differ between the groups (Time × Background, Time × Background × Emotion, all $F < 1.23$, all $p > .301$). These findings (i) demonstrate that the creation of negative versus neutral scenes had succeeded, (ii) show that no significant differences occurred between the groups in subjective ratings of valence and arousal, and (iii) underscore the necessity to analyze item and contextual memory in independent analyses, due to differently sized effects of negative versus neutral items in white backgrounds versus context images.

Negative items were recalled more often than neutral words in the baseline cued recall test (Emotion, $F_{1,154} = 8.59$, $p = .004$, $\eta_p^2 = .05$, Fig. S2). Also, the accuracy of items in background images was higher compared to items in white backgrounds (Background $F_{1,154} = 1857.74$, $p < .001$, $\eta_p^2 = .92$). These effects did not interact, however (Emotion × Background, $F_{1,154} = 0.03$, $p = .859$, $\eta_p^2 < .01$). Crucially, there were no significant differences in baseline cued recall accuracy between the groups (Time, Time × Emotion, Time × Background, Time × Emotion × Background, all $F < 0.69$, all $p > .560$).

Valence and arousal
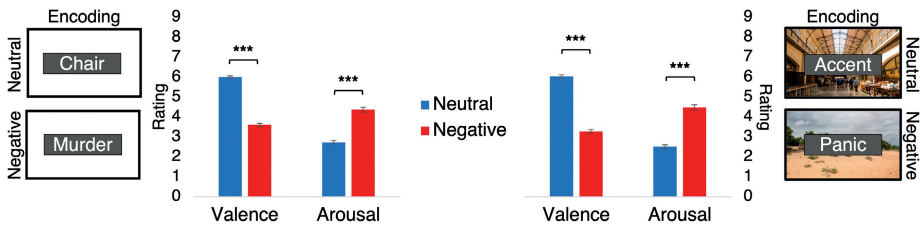
Fig. S1. Valence and arousal ratings during encoding on day 1. Ratings during neutral and negative item encoding on a white background (left) or a context image (right). ***$p < .001$. Error bars represent SEM.
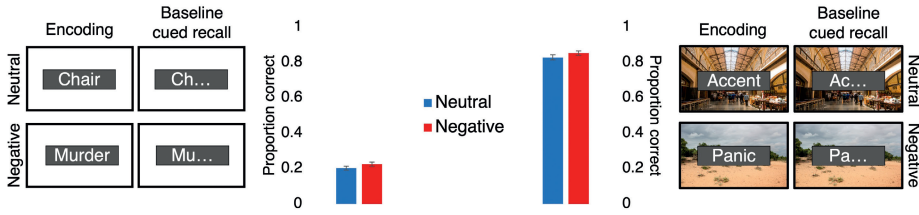
Baseline cued recall

Fig. S2. Assessment and results of baseline testing on day 1. Proportion of correct responses for neutral and negative items that were encoded and tested on a white background (left), or a context image (right). Error bars represent SEM.

2.2.   Primary results.

2.2.1.   Cued recall. The contrast analyses of cued recall difference scores (negative − neutral) did not reveal evidence for time-dependent changes in item memory or contextual dependency for negative versus neutral memories (item memory: C = -0.08, 95% CI [-0.30, 0.13], $F_{1,154}$ = 0.57, p = .450, η_p^2 < .01, contextual dependency: C = -0.07, 95% CI [-0.85, 0.71], $F_{1,151}$ = 0.03, p = .856, η_p^2 < .001, Fig. S3, S9).
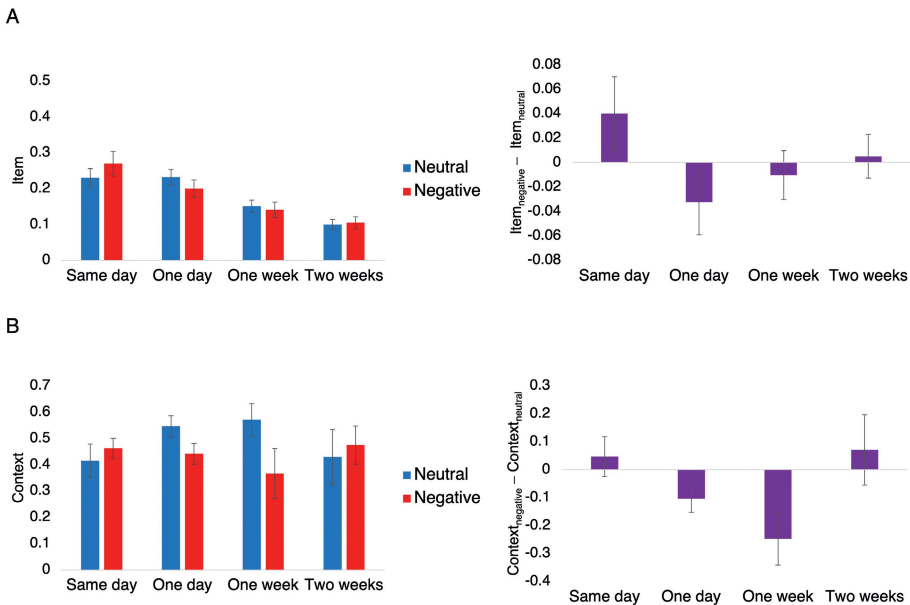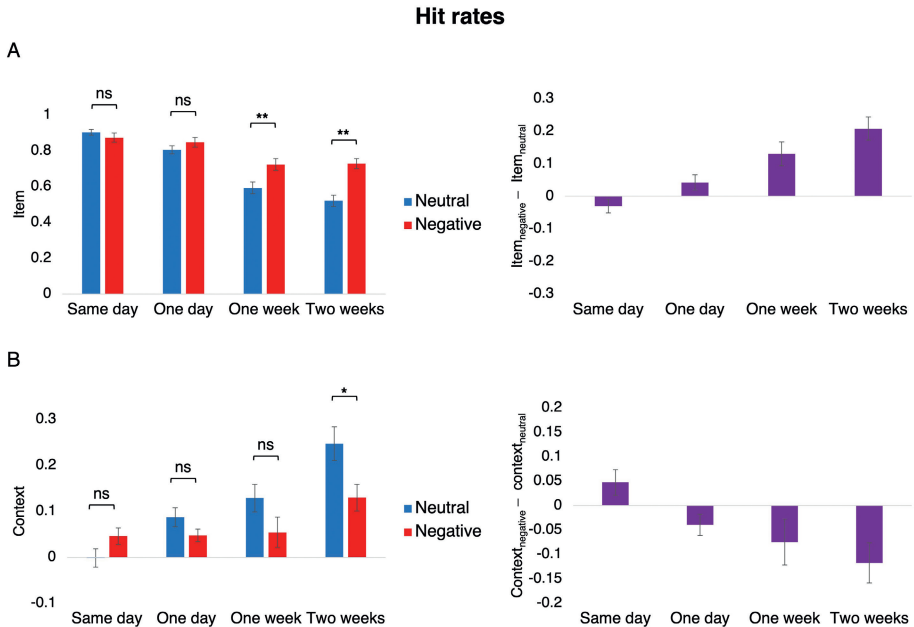
Cued recall

Fig. S3. (**A**) Item memory results displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). (**B**) Contextual dependency results displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). Error bars represent SEM.

2.2.2.   Hit rates. Analyses of hit rates showed increasingly stronger item memory and increasingly weaker contextual dependency over time for negative versus neutral memories (item: C = 0.80, 95% CI [0.54, 1.06], $F_{1,154}$ = 36.35, p < .001, η_p^2 = .19, context: C = -0.53, 95% CI [-0.84, -0.22], $F_{1,154}$ = 11.29, p = .001, η_p^2 = .07, Fig. S4, S8). These findings mirror the observations from the d-prime scores, thus ruling out that the findings are confounded by differences between the intact and rearranged context conditions at the first time point (i.e., the same day condition).

2

Fig. S4. (**A**) Item memory results displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). We observed a stronger retainment of negative items versus neutral items that became more pronounced with time. (**B**) Contextual dependency results displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right). Over time, contextual dependency of negative memories became increasingly smaller relative to neutral memories. * p < .05, ** p < .01, ns = not significant. Error bars represent SEM.

2.2.3.   False alarm rates. Another contrast analysis showed no time-dependent differences between negative and neutral memories for false alarms (C = 0.11, 95% CI [-0.12, 0.34], $F_{1,154}$ = 0.93, p = .336, η_p^2 < .01). These patterns did not differ between lures on white backgrounds and context images (C = -0.08, 95% CI [-0.39, 0.23], $F_{1,154}$ = 0.24, p = .628, η_p^2 < .01).

False alarm rates
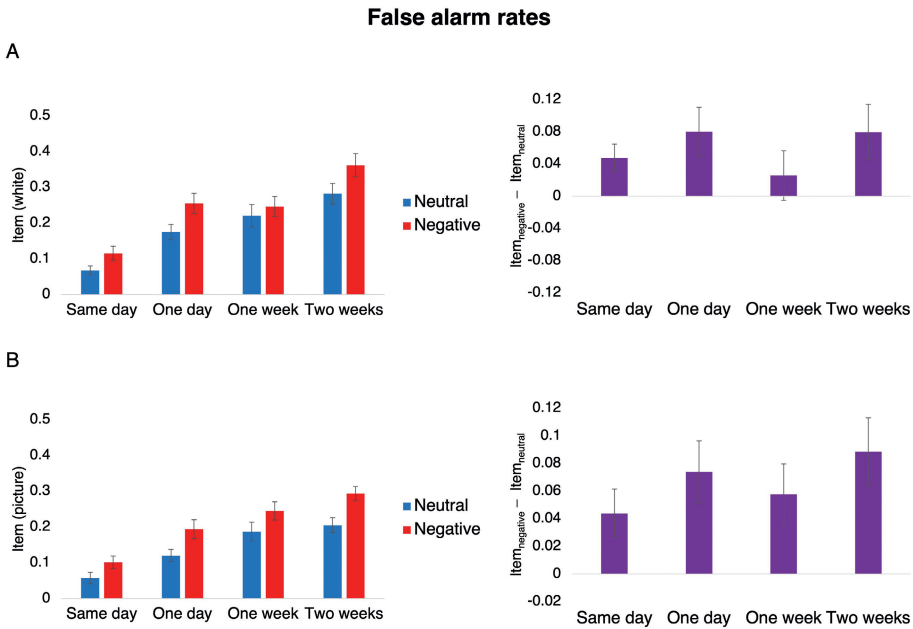
Fig. S5. Item memory results displayed for neutral and negative memories separately (left), and differences scores (negative − neutral, right) for lues in white backgrounds (**A**) and in context images (**B**). Error bars represent SEM.

2.3.   Secondary results.

2.3.1.   Cued recall. A time-dependent overall decrease of item memory was found ($F_{1,154}$ = 35.43 p < .001, η_p^2 = .19). Contextual dependency on the other hand did not show a reduction with the passage of time ($F_{1,151}$ = 0.11, p = .736, η_p^2 < .01, Fig. S3, S9). We also did not observe differences between negative memories and neutral memories in overall item memory accuracy ($F_{1,154}$ < 0.01, p = .961, η_p^2 < .01) or contextual dependency ($F_{1,151}$ = 1.77, p = .185, η_p^2 = .01).

2.3.2.   Hit rates. We observed an overall decrease of item memory, and increase in contextual dependency with time (item memory: $F_{1,154}$ = 85.59, p < .001, η_p^2 = .36, contextual dependency: $F_{1,154}$ = 36.76, p < .001, η_p^2 = .19, Fig. S4, S8). Also, item memory was larger, and contextual dependency smaller, for negative memories relative to neutral memories (item memory: $F_{1,154}$ = 34.86, p < .001, η_p^2 = .19, contextual dependency: $F_{1,154}$ = 6.85, p = .010, η_p^2 = .04).

2.3.3.   False alarm rates. We observed that with time false alarms generally increased ($F_{1,154}$ = 63.27, p < .001, η_p^2 = .29, Fig. S5). Furthermore, false alarms were overall more frequent for negative than neutral lures ($F_{1,154}$ = 45.06, p < .001, η_p^2 = .23). We found that none of these effects depended significantly on Background (picture versus white, all F < 3.02, all p > .084).

Together, the hit and false alarm rate data show exactly how recognition of negative items changes over time in comparison with neutral items. Even though false alarm rates are overall lower for neutral items than negative items, negative items show a retention of old item recognition relative to neutral items that becomes more pronounced over time. Consequently, neutral items show superior recognition at first, but with time the pattern reverses such that negative items are recognized better.
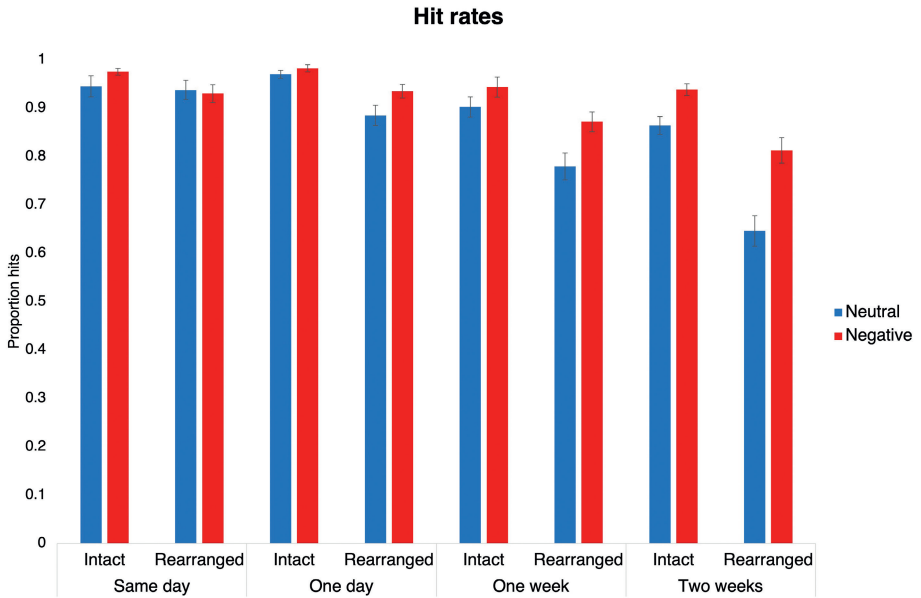
2

3.   Supplementary Figures
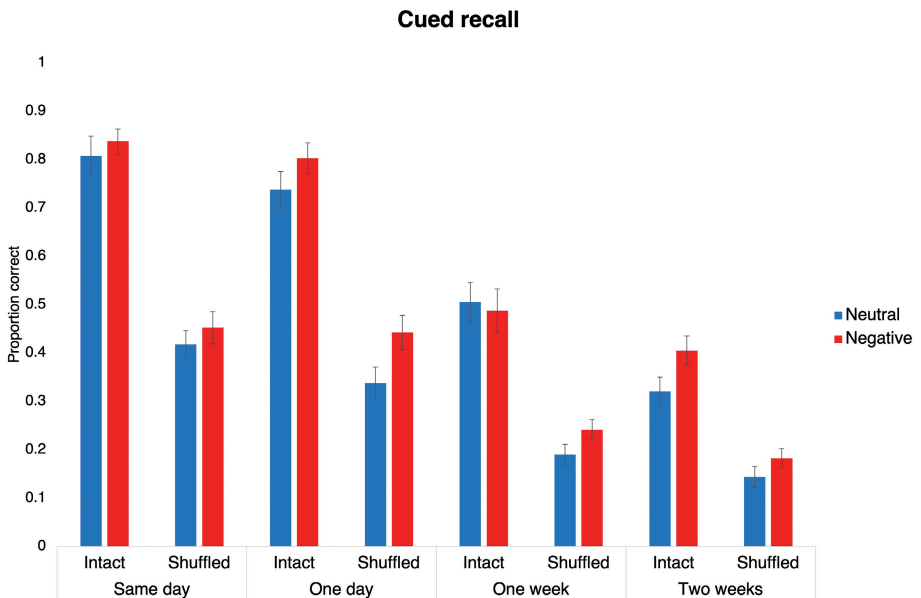
D-prime accuracy
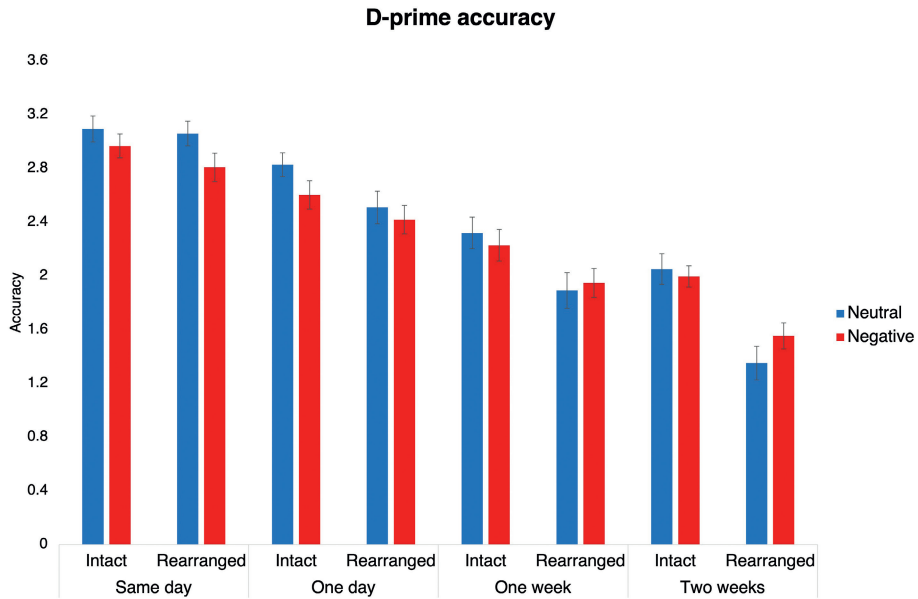
Fig. S6. Contextual dependency results displayed per group for neutral and negative items in intact and rearranged contexts. Error bars represent SEM.

Recollection
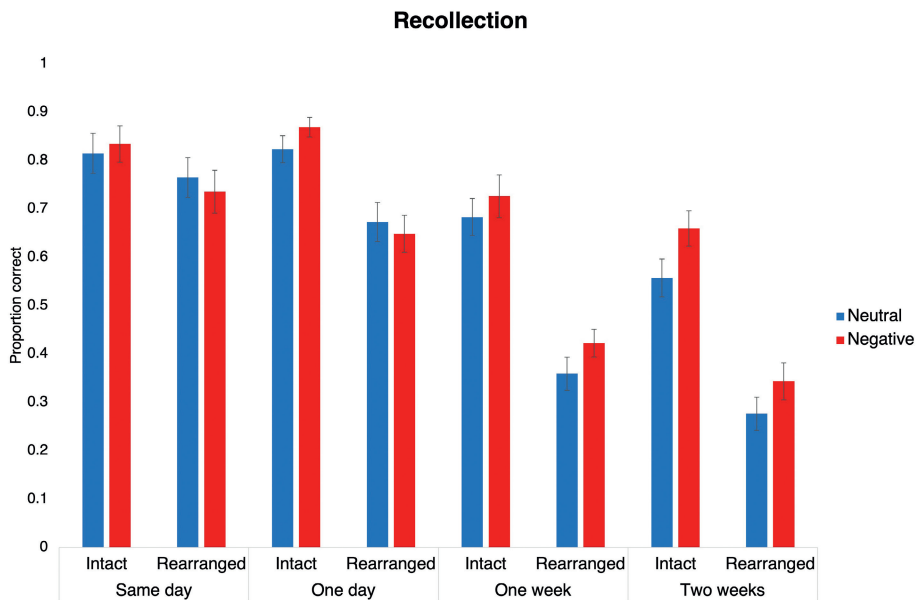
Fig. S7. Contextual dependency results displayed per group for neutral and negative items in intact and rearranged contexts. Error bars represent SEM.

Hit rates

Fig. S8. Contextual dependency results displayed per group for neutral and negative items in intact and rearranged contexts. Error bars represent SEM.

Cued recall

Fig. S9. Contextual dependency results displayed per group for neutral and negative items in intact and rearranged contexts. Error bars represent SEM.