POST-ENCODING INTERFERENCE AND THE PERSISTENCE OF EMOTIONAL MEMORY ENCODING

A Master's Thesis Presented to

the Faculty of the Department of Psychological Sciences

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

BRITTNEY M. BISHOP-CHRZANOWSKI

Dr. Jeffrey D. Johnson, Thesis Supervisor

MAY 2022

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

POST-ENCODING INTERFERENCE AND THE PERSISTENCE OF EMOTIONAL MEMORY ENCODING

presented by Brittney M. Bishop-Chrzanowski,

a candidate for the degree of Master of Arts,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Jeffrey Johnson

Dr. Nelson Cowan

Dr. Edgar Merkle

Dr. Ashley Curtis

Acknowledgements

I would like to express my gratitude to several individuals who helped throughout the course of my thesis project: First and foremost, my advisor, Dr. Jeffrey Johnson for his guidance and patience. I would also like to thank Dr. John Scofield and Kaitlyn Raith, a former undergraduate research assistant, for their preliminary efforts with this experiment. Finally, I am thankful to the members of my thesis committee for their invaluable input and feedback: Dr. Nelson Cowan, Dr. Ashley Curtis, and Dr. Ed Merkle.

Table of Contents

Abstract

Research on the emotional enhancement of episodic memory has largely focused on the inherent stimulus characteristics that might contribute to this effect. Less attention has been devoted to assessing the contributions of more general encoding processes that operate on both emotional and non-emotional stimuli. The current study focuses on how post-encoding elaboration contributes to the memory advantage for emotional information. Participants in an online study were presented with four categories of word stimuli to encode, based on levels of emotional valence (negative versus neutral) and arousal (high versus low). A math problem appeared after each word in attempt to terminate post-encoding elaborative processes, with the length of the delay between words and math problems (1500 versus 3500 ms) being the primary manipulation. Following a 48-hour delay, participants' memories for the words were tested using a "remember/know" procedure. We hypothesized that negative valence and high arousal would be associated with enhanced recollection-based memory performance relative to the neutral and low arousal conditions. Additionally, we predicted that the shorter word-math delay at encoding would mitigate the arousal-driven memory enhancement, consistent with several prior studies. Finally, we sought to test a novel hypothesis from a recent model of processing for emotionally-valenced stimuli (Bowen & Kensinger, 2017), in which the shorter delay was predicted to reduce the enhancement related to negative valence. The results of the current study provide partial support for these predictions (in particular, for arousal), but overall emotion-related effects were limited in size. Findings are discussed primarily in terms of potential limitations of our stimulus selection, experimental design choices, and general control of online studies of emotion and memory.

Keywords: Post-stimulus elaboration, valence, arousal, episodic memory

Post-encoding Interference and the Persistence of Emotional Memory Encoding

Episodic memory refers to the neurocognitive processes that allow for the encoding and conscious retrieval of spatio-temporally bound personal (i.e., autobiographical) events (Renoult et al., 2019; Tulving, 1972). Some episodic memories are more likely to be remembered than others, often due to recency or saliency; similarly, numerous experimental manipulations applied at the time of episodic memory encoding have been shown to affect later retrieval. Some of the most consistent of these manipulations involve the level (or depth) of elaborative processing (Craik & Lockhart, 1972; Craik & Tulving, 1975) and whether attention at encoding is fully available or divided by a secondary task (Craik et al., 1996; Naveh-Benjamin et al., 2000). Additionally, the nature of the to-be-remembered information can have a strong effect on memory. A clear example of this involves better memory for information that is emotionally-charged versus neutral – a phenomenon which has been demonstrated across a variety of laboratory stimuli including pictures (e.g., Bowen & Kensinger, 2017; Bradley et al., 1992; Kensinger et al., 2011), words (e.g., Brainerd, 2018; Brainerd & Bookbinder, 2019; Doerksen & Shimamura, 2001), and video clips (e.g., Samide et al., 2019). The current study is designed to better understand the encoding processes that might contribute to this memory advantage for emotional information.

Considering that various processes contribute to memory encoding, one might gain further understanding by segregating processes into distinct levels of operation (for review, see Cohen et al., 2015). At an early stage are pre-stimulus processes that could be used to prepare for encoding the upcoming stimulus (e.g., Addante et al., 2015;

Otten et al., 2006). Effects broadly related to vigilance or attention could be included in this category, given that they are ongoing rather than elicited by individual stimuli (e.g., Naveh-Benjamin et al., 2003). Next are stimulus-specific operations thought to be elicited by either individual stimuli or subject to inter-trial experimental changes. Examples of these operations include processes related to inherent stimulus characteristics (such as meaning or relevance; see Madan, 2021), the amount and nature of elaboration cued directly by the stimulus, or other manipulations such as presentation duration and depth of encoding (e.g., Leiker & Johnson, 2014; Rugg et al., 2000; Vilberg & Rugg, 2009). Finally, several post-stimulus processes have been shown to affect later retrieval, from continued elaboration immediately following encoding (e.g., Ben-Yakov & Dudai, 2011; Ben-Yakov et al., 2013; Hulbert et al., 2016; for a review of directed forgetting see Hulbert et al., 2018) to the reactivation of episodic information during subsequent periods of wakeful rest (e.g., Staresina et al., 2013; Tambini et al., 2010) and sleep (e.g., Diekelmann et al., 2012; Dueker et al, 2013; for recent meta-analysis, see Hu et al., 2020). In one example, Ben-Yakov and Dudai (2011) found that hippocampal activity immediately following the offset of a video stimulus differed from activity seen during stimulus presentation and was distinctly predictive of retrieval success.

In the context of emotional memory, the vast majority of research has focused on stimulus characteristics and the concomitant encoding processes thought to enhance or bias memory formation (e.g., Mather & Sutherland, 2011; for reviews, see LaBar & Cabeza, 2006; Talmi, 2013; Yonelinas & Ritchey, 2015;). Much less attention has been devoted to assessing the contributions of post-stimulus processes (cf. Bradley et al., 1992; Christianson, 1992; Hulse et al., 2007; see Bowen et al., 2017, for a recent discussion of this issue). To elucidate these contributions, it is first necessary to un-confound two dimensions of emotion: *valence* and *arousal*. Valence can be defined as the degree of positivity or negativity (i.e., pleasantness or unpleasantness) associated with a given stimulus, whereas arousal refers to the amount of excitement or calmness it induces (Lang et al., 1993; Russell, 1980).

Supporting this characterization, Kensinger and Corkin (2004) used functional magnetic resonance imaging (fMRI) in a subsequent memory paradigm (e.g., Brewer et al., 1998; Wagner et al., 1998; for review, see Paller & Wagner, 2002) to demonstrate separate neural correlates of valence and arousal. While the hippocampus exhibited subsequent memory effects for both dimensions, encoding-related activity in inferior prefrontal cortex (PFC) for the valence dimension was distinct from that in the amygdala for arousal (all left-lateralized). In a companion behavioral experiment, Kensinger and Corkin divided participants' attention with a secondary task at encoding, showing that the valence effect was disproportionately reduced in comparison to the arousal effect, which remained intact. Together, these findings suggest that, whereas the arousal enhancement for memory reflects relatively automatic encoding processes, valence relies on controlled (i.e., self-generated) elaboration that continues beyond initial stimulus presentation (Kensinger & Corkin, 2004).

The distinction noted above between automatic and controlled processing aligns with Christianson's (1992) long-standing theory of emotional memory enhancement, in which both pre-attentive processing and post-stimulus elaboration are contributing factors. Christianson's theory, however, is thought to be more applicable to the dimension of arousal than valence (for recent discussion, see Bowen et al., 2017). Indeed, many studies directly examining these processes involved negatively-valenced stimuli that were also highly arousing (compared with low-arousal, neutral stimuli), a confound which leaves open the possibility that arousal was driving any enhancement to retrieval (e.g., Harris & Pashler, 2005; Heuer & Reisberg, 1990; Migita et al., 2011; cf. Libkuman et al., 2004).

In an extension of the aforementioned work, Bowen and colleagues (2017) recently proposed a model focused explicitly on the reactivation of negatively-valenced memories. This Negative Emotional Valence Enhances Recapitulation (NEVER) model is based on several influential theories of episodic memory which state that the operations and stimulus characteristics activated during encoding are reactivated (or recapitulated) at the time of retrieval (Alvarez & Squire, 1994; McClelland et al., 1995; Rolls, 2000). By this account, encoding elicits the activation of sensory details that can later serve as the basis for detail-based retrieval (i.e., recollection; Tulving & Thomson, 1973; Morris et al., 1977). Moreover, the amount of overlap between patterns of brain activity at encoding and again at retrieval has been correlated with retrieval performance (e.g., Johnson & Rugg, 2007; Khan et al., 2004; Wheeler et al., 2000; for reviews, see Danker & Anderson, 2010; Rugg et al., 2008). The NEVER model posits that such encoding-retrieval overlap is more likely to occur for stimuli of negative valence than for stimuli that is either neutral or positive (e.g., Bowen & Kensinger, 2017; Ford et al., 2014; Kensinger & Schacter, 2008; Mickley Steinmetz & Kensinger, 2009; for review, see Kensinger & Ford, 2020). In particular, the model suggests that the emotion enhancement associated with negative valence begins with sensory-focused processing at encoding, is further driven by recapitulation at retrieval, and can also occur during intermittent offline periods (e.g., immediately following encoding and during rest/sleep).

Current Experiment

The motivation for the current experiment was to further elucidate the role of post-stimulus elaboration in the enhancement for emotional memory encoding. As outlined above, the evidence for post-stimulus elaboration is inconsistent, with some reports of it being related to arousal (e.g., Bradley et al., 1992; Christianson, 1992) and, more recently, others proposing that it also operates for negatively-valenced stimuli (Bowen et al., 2017). Here, we used word stimuli (Warriner et al., 2013) selected to be orthogonal with respect to valence (negative versus neutral) and arousal (high versus low) to directly test for these differences. Under the assumption that post-stimulus processing should continue for negative as well as highly arousing words, a secondary, interfering task consisting of math problems was inserted immediately after each encoding trial (for similar approaches, see Libkuman et al., 2004; Migita et al., 2011). Our secondary task manipulation involved the inter-stimulus interval between the offset of the to-be-remembered word and the onset of the math problem (1500 versus 3500 ms), with the shorter duration predicted to cease (or decrease) elaborative processing.

Following the encoding phase, we tested participants' memories with a "remember/know" procedure (described in greater detail below) that allowed us to distinguish between recollection- and familiarity-based memory (for review, see Yonelinas, 2002). Importantly, the retrieval task was administered approximately 48 hours after encoding, as prior studies and our preliminary findings suggest that the presence of the emotional memory enhancement is inconsistent with shorter delays (e.g., Kleinsmith & Kaplan, 1963; Sharot & Phelps, 2004; Sharot & Yonelinas, 2008).

The current study is based on two hypotheses regarding performance on the retrieval task. First, we predicted that both negative valence and high arousal would be associated with better retrieval performance than their neutral and low arousal counterparts. We particularly expected these effects for recollection-based responses, as previous research has shown minimal effects of manipulations on familiarity-based responses (e.g., Bowen & Kensinger, 2017; Kensinger & Corkin, 2003; Naveh-Benjamin and Kilb, 2012). Secondly, and consistent with the NEVER model (Bowen et al., 2017), we anticipated that the secondary task manipulation at encoding would mitigate the valence enhancement, in addition to the expected arousal enhancement (Bradley et al., 1992; Christianson, 1992). Specifically, we anticipated that the emotion-related effect following the short (1500 ms) delay interval would be smaller than that following the long (3500 ms) interval.

Methods

Participants

Sixty-eight participants were recruited from the University of Missouri (MU) community through an undergraduate psychology participant pool and online advertisements. Participants were compensated with either research course credit or $5 for each of the two parts (the *study phase* and *test phase*; each lasting less than 30 minutes). All participants self-reported to be right-handed, native-English speaking, and between 18 and 30 years of age. Informed consent was obtained in accordance with the MU Institutional Review Board.

Sixteen participants were excluded from all analyses because they completed only one part of the experiment. Data for the remaining participants were then screened using a multi-step process that identified outliers in terms of high rates of missing responses (for either the study or test phase), low performance on the math problems in the study phase, and low overall memory performance on the test [based on corrected recognition (Pr: hit rate minus false alarm rate); Snodgrass & Corwin, 1988]. One participant was identified for failing to respond on 87% of the test trials, and another participant performed poorly on the math problems (only 29% correct). Three additional participants had overall memory scores (Pr) that were at or below zero, which was taken as an indication of misunderstanding of or noncompliance with the task instructions, and thus their data were excluded.

The final sample of 47 participants had a mean age of 21 years (SD = 3). Thirty-six of the participants self-identified as women, eight as men, and three as non-binary.

Additionally, 85% of the sample reported being White/Caucasian, 4% Black/African American, 9% Asian/Pacific Islander, and 2% More than one race.

Materials

Word Stimuli

Word stimuli were drawn from the Warriner, Kuperman, and Brysbaert (2013) database, which includes 13,915 English words with ratings of emotional valence, arousal, and dominance. The rating scales ranged from 1 (negative valence or “unhappy”; low arousal or “calm”; low dominance or “controlled”) to 9 (positive valence or “happy”; high arousal or “excited”; high dominance or “in control”), with the middle of the scale at 5 (“neutral”). For the current study, words were selected from each of four conditions: negative valence, high arousal; negative valence, low arousal; neutral valence, high arousal; and neutral valence, low arousal. The dominance ratings were ignored.

To identify stimuli that fit into the four conditions noted above, the words in the original database were initially sorted. For the arousal ratings, cutoffs were placed at the 18th and 82nd percentiles, providing words that were low or high on this dimension. For valence, a cutoff was placed at the 18th percentile to identify words with negative ratings, but words of neutral valence were selected by centering the percentile range around the neutral rating 5 (from the 41st to 59th percentile). These cutoffs were adjusted to be as narrow as possible, while also being equally sized across conditions and providing enough stimuli for each of the corresponding valence × arousal combinations. Words were further excluded if they exhibited substantial gender differences in ratings (mean difference > 1 in either direction; also provided by Warriner et al., 2013) or were lexically similar to other words (e.g., either *depressed* or *depression* was included, but not both). The numbers of words surviving these criteria were as follows: negative valence, high arousal (n = 494); negative valence, low arousal (67); neutral valence, high arousal (47); and neutral valence, low arousal (207). From these remaining words, 40 were randomly selected from each valence × arousal combination to serve as the experimental stimuli. Descriptive statistics for the valence and arousal ratings of final words included in this experiment are provided in Table 1.

Table 1. Valence and arousal ratings of word stimuli selected for each emotion condition

	Valence			Arousal		
Stimulus condition	M	SD	Range	M	SD	Range
Neutral valence, low arousal	5.04	0.10	4.84–5.17	3.02	0.28	2.29–3.36
Neutral valence, high arousal	5.03	0.12	4.84–5.18	5.48	0.31	5.05–6.52
Negative valence, low arousal	3.20	0.46	2.00–3.80	2.97	0.38	1.67–3.38
Negative valence, high arousal	2.78	0.59	1.54–3.73	5.76	0.57	5.10–7.24

Math Stimuli

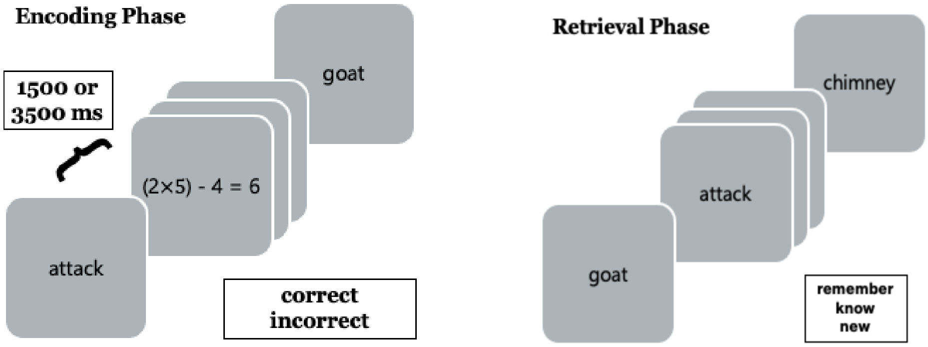
Math stimuli, which were presented only during the study phase, consisted of two-step problems followed by either a correct or incorrect answer [e.g., $(6 * 4) + 9 = 35$ and $(2 * 5) - 4 = 6$]. The rules for generating these problems were as follows: (1) three single-digit numbers ranging between 1 and 9 were presented on the left side of the equation; (2) a multiplication sign (asterisk, *) separated the first two numbers, with surrounding parentheses indicating that this part of the problem should be computed first; (3) an addition or subtraction sign preceded the third number; (4) incorrect answers on the right side of the equation were chosen by randomly selecting a number ±2 from the correct answer; and (5) answers were restricted to positive values. These rules were used to generate a set of 2,762 problems, from which 80 were randomly selected (equal numbers with correct and incorrect answers) to be presented to each participant.

All words, math problems, and other text stimuli (e.g., fixation markers) were presented in white 16-point Arial font against a gray background. Words were also presented in lowercase. Because the experiment was run online, there was no control over the display size of each participant's computer, viewing distance, or the web browser used to complete the experiment.

Procedure

The experiment was created in PsychoPy (v2021.2.3; https://www.psychopy.org; Peirce et al., 2019) and administered online via Pavlovia (https://pavlovia.org), an open source platform for launching PsychoPy experiments. After giving informed consent and responding to a brief demographics survey (both completed via Qualtrics; Provo, UT), participants were provided a link to the study phase, which lasted about 20-30 minutes. Following the study phase, participants were provided a link to return to the experiment in about 48 hours to complete the second part (the test phase). A general schematic of the study and test phases is provided in Figure 1. Although participants were encouraged to complete the second part as close as possible to the 48-hour mark, the test phase was always accessible in order to provide some flexibility in timing. For our final sample (N = 47), the test phase was completed at an average of 49.4 hours (SD = 16.1) after the study phase. After the test phase, which also lasted about 20-30 minutes, participants were debriefed about the nature of the experiment. They were then taken to an additional, optional questionnaire (11 questions) about the quality and duration of their sleep the night before. The sleep questionnaire served to pilot methods for future studies potentially linking memory and sleep, and thus the data were not analyzed for purposes of the current study. Additionally, because these questions took place at the end of the experiment, they would not have influenced any of the findings reported here.

Figure 1. Schematic of stimulus presentation and response options in the study (left) and test (right) phases



Study Phase

The study phase began with instructions and practice on the task and procedures. The study phase consisted of 80 words (20 from each valence × arousal condition) that were presented one at a time, and each word was followed by a math problem. The words were organized into four blocks consisting of 20 words each (equally drawn from the four conditions), with self-paced breaks dividing the blocks. There were two types of study blocks that were made explicitly clear to participants in the instructions and practice. The different blocks were distinguished by whether there was a short (1500 ms) versus long (3500 ms) inter-stimulus interval (ISI) between the offset of the word stimulus and the onset of the following math problem. The sequence of centrally-presented stimuli on all trials began with a red fixation cross (+) for 500 ms, followed by a word for 750 ms. The ISIs during the short and long delay blocks then consisted of presentation of either one or three tilde characters (∼ or ∼∼∼), respectively. The math problem was displayed for 2500 ms following the ISI, and finally, a white fixation cross appeared for 3500 or 1500 ms for the short vs. long ISI trials, respectively, to equate the overall trial length across block types (at 8,750 ms). The two block types were administered in an ABBA order, with the first type counterbalanced across participants.

There was no response requirement for the study words; participants were instead instructed from the outset to intentionally encode the words into memory for a later test. For the math problems, participants were to indicate whether the equation was correct or incorrect by respectively pressing the “J” and “K” keys on the keyboard with their right index and middle fingers. The response mapping was also displayed below each math problem for the entire presentation duration. The instructions encouraged participants to make their responses within the 2500 ms that the math problem was presented.

Test Phase

The test phase consisted of presentation of the 80 words from the study phase (hereafter, *old* words), which were randomly intermixed with the remaining 80 words from the stimulus pool (*new* words; 20 from each valence × arousal condition). Each test trial consisted of central presentation of a white fixation cross for 500 ms and then a word for 3000 ms. Participants were instructed to respond to each word according to the standard *remember/know/new* procedure (Tulving, 1985; Rajaram, 1993). A *remember* response was to be used when any specific details surrounding a word's presentation during the study phase could be retrieved; a *know* was to indicate a feeling of familiarity that the word was presented during the study phase, but in the absence of retrieving any specific details; and words not presented during the study phase were to be designated as *new*. These respective responses were made by pressing the "J", "K", and "L" keys with the right index, middle, and ring fingers. The response mappings were displayed below the fixation cross and word stimuli continuously throughout the test phase. Participants were instructed to attempt to respond within the 3000-ms presentation window of each word, before it disappeared.

Results

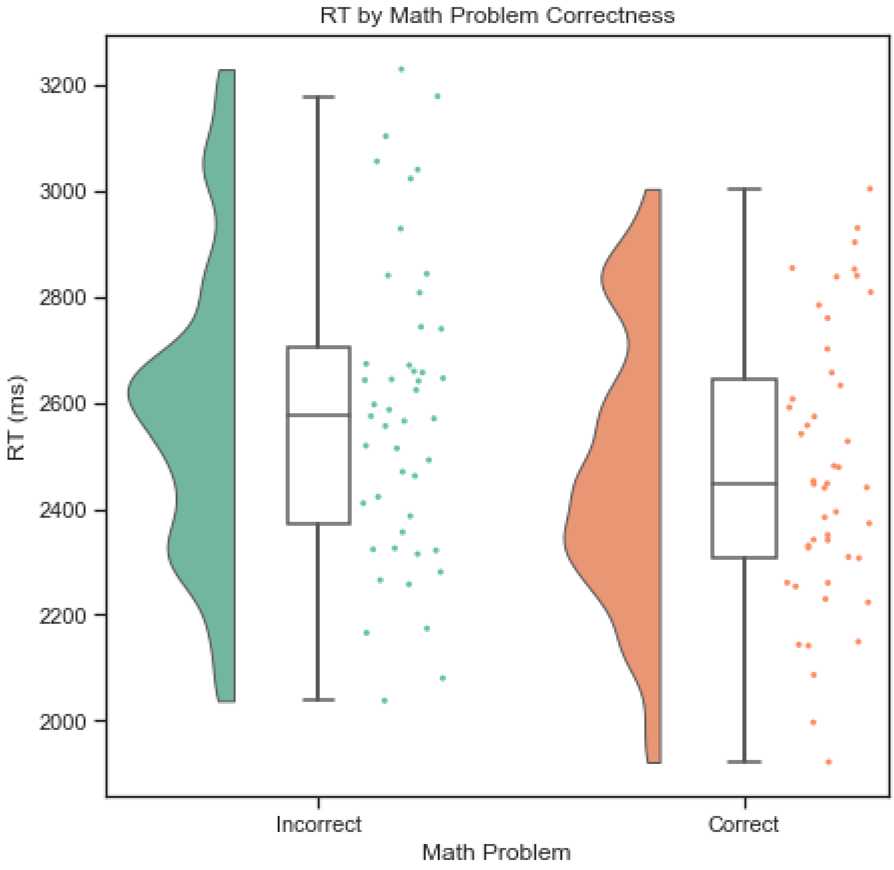
Study Phase

Following the participant screening procedures (described in Methods), the remaining participants (N = 47) failed to respond to 5.66% of the math problems in the study phase (SD = 5.95%; range = 0 to 26.25%). These missing responses were treated as incorrect for subsequent analyses of the study phase data.

To assess performance on the math problems in the study phase, we analyzed both the accuracy participants' responses in labeling the math problems as correct versus incorrect and the associated response times (RTs). Overall, participants responded correctly to 81.81% (SD = 9.52%) of the math problems, which was significantly greater than chance (0.5, t(46) = 22.90, p = 4.32e-27, d = 3.34, one-tailed). To assess whether accurate responding differed with respect to the status of the math equations – i.e., whether the solutions provided were correct versus incorrect – the data were also separated according to this factor. The resulting correct response rates were 80.85% (SD = 9.99%) and 82.76% (SD = 11.69%) for correct and incorrect equations, respectively. A paired-samples t-test indicated that these means were statistically indistinguishable (t(46) = 1.25, p = 0.218, d = 0.18).

Finally, we analyzed the math problem RTs, which are summarized in Figure 2. Overall, the mean RT was 2528 ms (SD = 261 ms), irrespective of the correct or incorrect status of the equations and whether the participant's response was correct or incorrect. Separating the RTs according to correct/incorrect status indicated that responses to correct equations (M = 2474 ms, SD = 261) were significantly faster than those to incorrect equations (M = 2583 ms, SD = 284; t(46) = 4.93, p = 1.1e-5, d = 0.40). Note that because performance on the math problems was rather high, several participants did not have enough incorrect responses to compute meaningful descriptives of the corresponding RTs.

Figure 2. Study phase response times (RTs) according to the correct/incorrect status of math problems

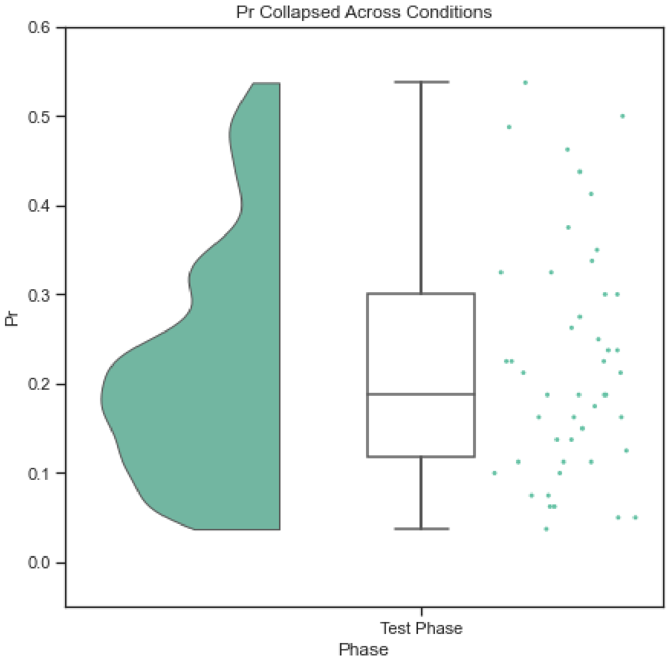


RTs (reported in milliseconds) are segregated according to whether the solutions provided for math problems were correct or incorrect. A density plot and boxplot are provided for each condition, along with the means for individual subjects. For the boxplots, the median is indicated by a horizontal line, the box represents the interquartile range (IQR; 25th to 75th percentiles), and the whiskers extend to 1.5×IQR.

Test Phase

As described in the Methods, participants were excluded from analysis if they had high missing response rates or low overall memory performance (Pr: hit rate minus false alarm rate) in the test phase. The missing response rate for the remaining participants was 1.06% (SD = 3.18%; range = 0 to 20.63%). For the overall Pr, which notably was collapsed across remember/know responses and across all of experimental manipulations (valence, arousal, and math delay), a mean of 0.22 (SD = 0.13; range = 0.04 to 0.54) was observed. The Pr measure is summarized for the group and plotted for individual participants in Figure 3. Although some participants had overall Pr values that were close to but just above zero, their data were still included in all of the analyses due to our expectations that the 48-hour study-test delay and secondary math task at study would might limit memory performance.

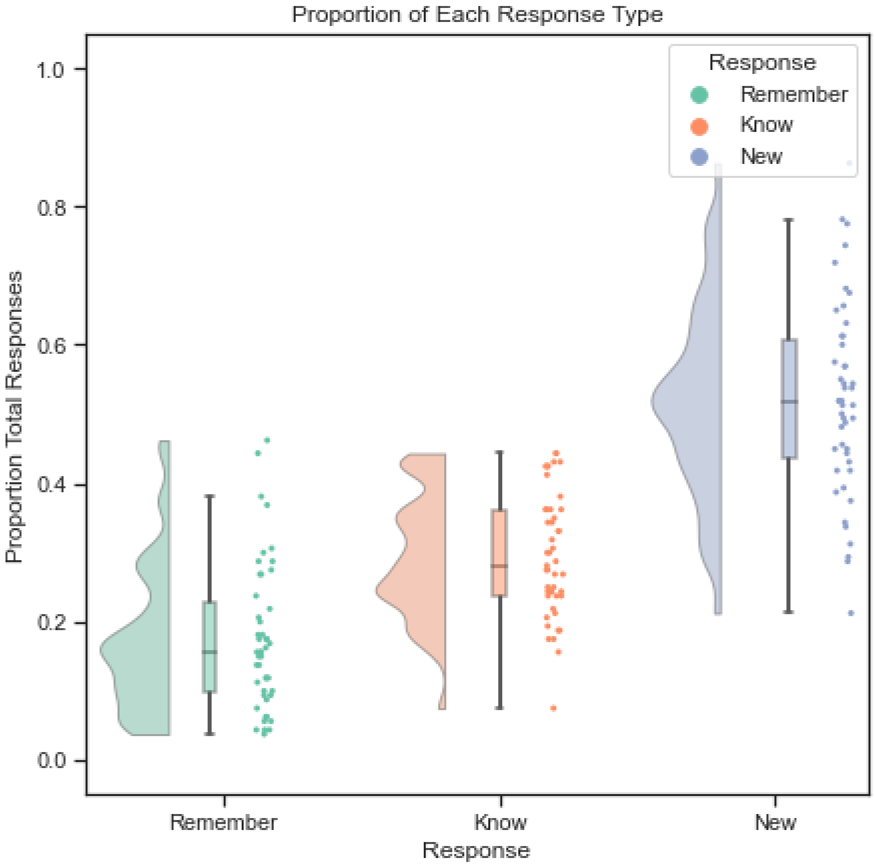
Figure 3. Overall memory performance (Pr = hit rate minus false alarm rate) during the test phase



Performance (Pr) was computed by combining the remember and know responses and is shown collapsed over all of the experimental manipulations (valence, arousal, and delay). For further plot details, see the caption of Figure 2.

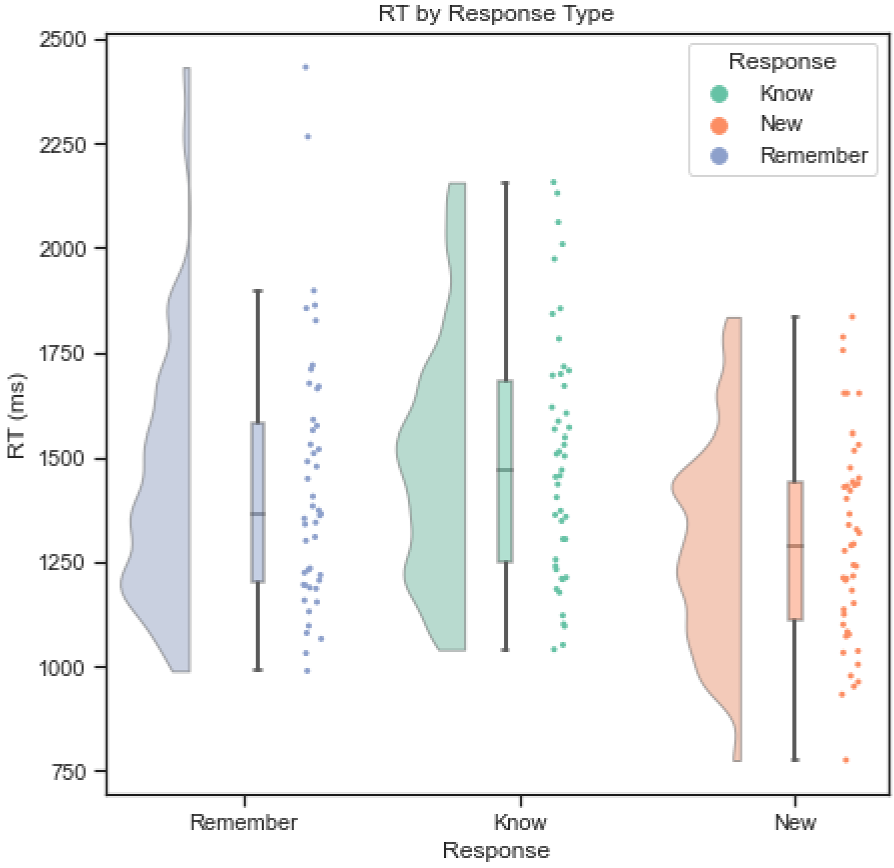
Once the data were screened, we next focused on participants' responding according to the remember/know/new procedure of the test phase. The proportions of each response type and the corresponding RTs, regardless of any of the experimental manipulations and whether responding was correct versus incorrect, are summarized in Figures 4 and 5, respectively. As can be seen, new responses were the most frequent (M = 0.52, SD = 0.22), followed by know (M = 0.29, SD = 0.14) and then remember responses (M = 0.17, SD = 0.16). Regarding the RTs, new responses were the fastest (M = 1285 ms, SD = 504 ms) and know responses were the slowest (M = 1485 ms, SD = 506 ms), with remember responses falling in between (M = 1379 ms, SD = 499 ms). For the remainder of the test phase analyses, we focus on the response rates and other measures computed from them (see below), since the low response rates for some of the conditions make it difficult to compute meaningful RT descriptives.

Figure 4. Overall proportions of remember, know, and new responses during the test phase



These proportions are collapsed over the manipulations of valence, arousal, delay, and old/new status. For further plot details, see the caption of Figure 2.

Figure 5. Overall response times (RTs) for remember, know, and new responses during the test phase



These RTs are collapsed over the manipulations of valence, arousal, delay, and old/new status. For further plot details, see the caption of Figure 2.

Next, to begin to assess memory performance, the remember, know, and new response proportions were separated according to the old/new status of test items and the valence and arousal manipulations (the math delay is introduced below). Although we were ultimately interested in response differences across the valence and arousal conditions, we computed, as described later, additional measures based on the proportion data to account for the non-independence of remember and know responses (Yonelinas et al., 1998). The proportion data are summarized in Table 2. For remember responses, old items (M = 0.25, SD = 0.14) were endorsed at a higher rate than new items (M = 0.10, SD = 0.09; t(46) = 9.82, p = 7.2e-13, d = 1.28). The same pattern was evident for know responses (Ms = 0.33 and 0.26, SDs = 0.10 and 0.09, respectively; t(46) = 5.30, p = 3.0e-6, d = 0.71). As expected, the opposite difference was observed for new responses, such that new items were endorsed more frequently (M = 0.63, SD = 0.15) than old items (M = 0.41, SD = 0.16; t(46) = 11.17, p = 1.1e-14, d = 1.41).

Table 2. Proportions (M and SD) of remember (R), know (K), and new (N) responses according to emotion condition and old/new status

	Old			New		
Stimulus condition	R	K	N	R	K	N
Neutral valence, low arousal	0.17 (0.14)	0.31 (0.15)	0.51 (0.19)	0.06 (0.07)	0.22 (0.11)	0.72 (0.15)
Neutral valence, high arousal	0.25 (0.17)	0.31 (0.13)	0.43 (0.19)	0.08 (0.10)	0.22 (0.11)	0.69 (0.16)
Negative valence, low arousal	0.25 (0.15)	0.36 (0.13)	0.39 (0.18)	0.09 (0.10)	0.28 (0.13)	0.61 (0.17)
Negative valence, high arousal	0.33 (0.20)	0.33 (0.14)	0.32 (0.19)	0.16 (0.14)	0.32 (0.14)	0.50 (0.20)

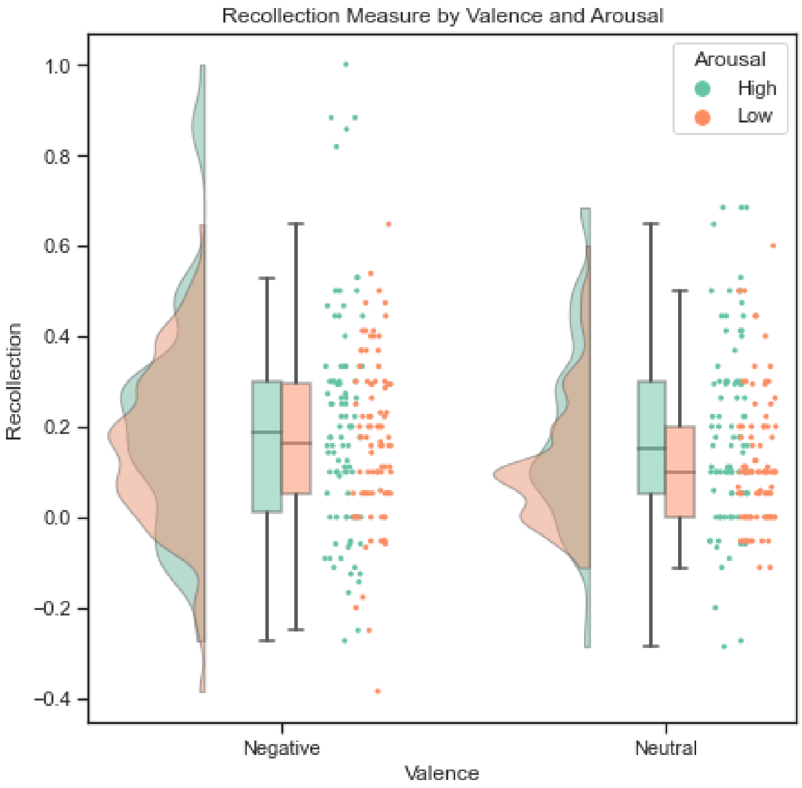
Central to our main hypotheses was first establishing the effects of emotional valence (negative versus neutral) and arousal (high versus low) on test memory performance. To obtain a measure of recollection based on remember responses, while also accounting for the overall tendency that participants had to make such responses, we applied a standard correction of these data (Yonelinas et al., 1998; also see

Kensinger & Corkin, 2003). Specifically, the proportion of new items designated as remembered (R_{new}) was subtracted from the proportion of remembered old items (R_{old}), and then that difference was divided by one minus R_{new}. The measures of recollection for each valence × arousal combination are summarized in Table 3 and Figure 6.

Table 3. Recollection and familiarity measures (M and SD) according to the emotion and delay (overall and split) conditions

	Overall		Long delay		Short delay	
Stimulus condition	Rec.	Fam.	Rec.	Fam.	Rec.	Fam.
Neutral valence, low arousal	0.12 (0.13)	0.51 (0.60)	0.12 (0.15)	0.47 (0.84)	0.11 (0.13)	0.46 (0.73)
Neutral valence, high arousal	0.18 (0.17)	0.56 (0.49)	0.20 (0.21)	0.54 (0.61)	0.17 (0.20)	0.49 (0.77)
Negative valence, low arousal	0.17 (0.15)	0.54 (0.68)	0.15 (0.20)	0.51 (0.82)	0.19 (0.16)	0.53 (0.73)
Negative valence, high arousal	0.20 (0.22)	0.35 (0.74)	0.19 (0.26)	0.32 (0.94)	0.21 (0.23)	0.42 (0.93)

Figure 6. Recollection measures for the test phase, according to valence and arousal



Recollection was computed as $(R_{old}–R_{new})/(1–R_{new})$ to take into consideration the overall tendency that participants had to make remember responses. For further plot details, see the caption of Figure 2.

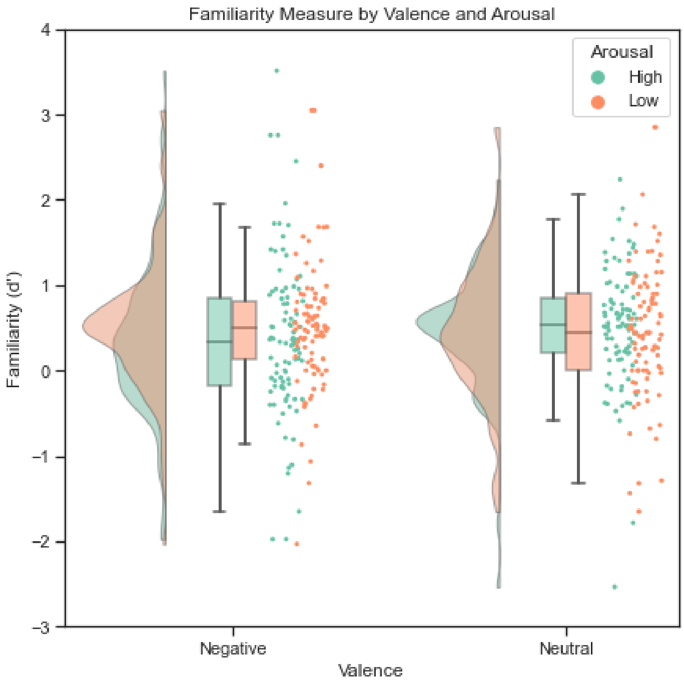
For recollection, participants had an overall mean score of 0.17 (SD = 0.12), which was significantly above the chance level of zero (t(46) = 9.59, p = 7.5e-13, d = 1.40). To test for emotion-related differences, a repeated-measures two-way ANOVA with factors of valence (negative versus neutral) and arousal (high versus low) was conducted. The ANOVA gave rise to a main effect of arousal (F(1, 46) = 6.69, p = 0.013, η_p^2 = 0.13), such that words rated as being highly arousing (M = 0.19, SD = 0.20) were recollected at a higher rate than those rated low on arousal (M = 0.14, SD = 0.14). Although the effect of valence did not reach significance (F(1, 46) = 2.96, p = 0.092, η_p^2 =

0.06), the difference was in the predicted direction, whereby recollection was slightly higher for negative (M = 0.19, SD = 0.19) than for neutral words (M = 0.15, SD = 0.15). The ANOVA also revealed that the interaction between valence and arousal was non-significant ($F(1, 46) = 0.495$, $p = 0.528$, $\eta_p^2 = 0.009$).

While measures of familiarity can also be obtained from the remember/know/new procedure, it is again important to correct the data accordingly. Here, the correction is based on participants having been instructed to respond know only if an item is not remembered. Therefore, a two-step procedure was used to compute familiarity (also see Yonelinas et al., 1998; also see Kensinger & Corkin, 2003). First, familiarity was computed separately for old items (F_{old}) as the proportion of old items designated as known (K_{old}) divided by one minus R_{old}, and for new items (F_{new}) as the proportion of new items designated as known (K_{new}) divided by one minus R_{new}. Then, these values were z-transformed, based on a normal distribution with M = 0 and SD = 1, and the corresponding F_{new} value was subtracted from that for F_{old} to give a difference in familiarity (hereafter, $F_{d'}$). It is important to note that, because F_{old} and F_{new} can take on values of 1 and 0, and this becomes increasingly common when the data are further separated according to the valence, arousal, and math delay manipulations, the z-transformation can produce infinite values. To solve this problem, values of 1 and 0 were respectively replaced with 0.99 and 0.01 (cf. Snodgrass & Corwin, 1988). The resulting familiarity measures for the valence × arousal combination are detailed in Table 3 and Figure 7.

As was the case with recollection, the overall familiarity score (M = 0.49, SD = 0.63) was also significantly above chance (vs. 0: t(46) = 9.28, p = 2.1e-12, d = 1.35). To identify any effects of valence or arousal on familiarity, these scores were submitted to a two-way repeated-measures ANOVA as was done for recollection. However, this ANOVA revealed no significant effects (all ps > 0.20, all $\eta_p^2 < 0.04$). Closer examination of the data indicated that the valence and arousal differences were actually in the opposite directions of our predictions, with $F_{d'}$ higher for low (M = 0.52, SD = 0.64) compared to high arousal (M = 0.46, SD = 0.63) and higher for neutral (M = 0.53, SD = 0.55) than negative valence (M = 0.45, SD = 0.71).

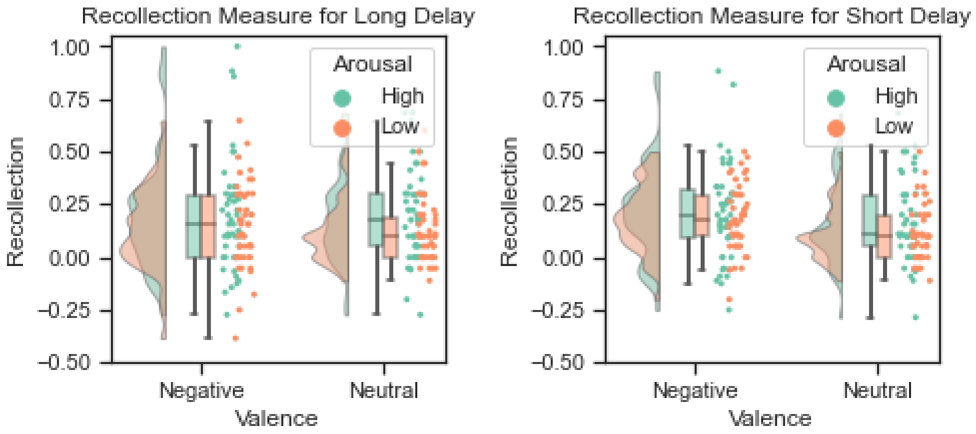
Figure 7. Familiarity measures for the test phase, according to valence and arousal

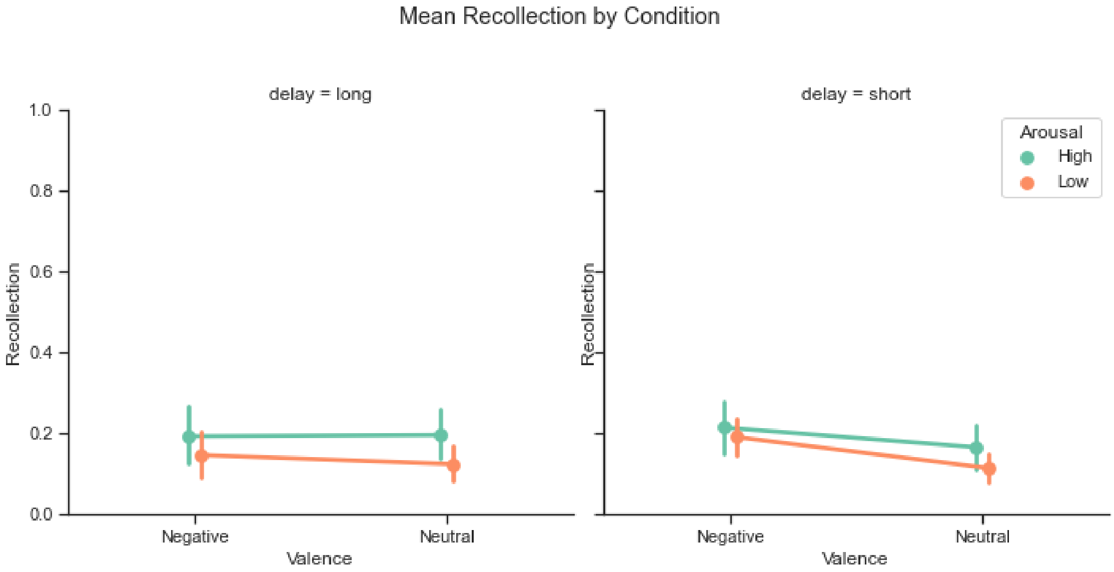


Familiarity is reported as a d' (difference) score based on the z-transformed values of F_{old} [$K_{old}/(1–R_{old})$] and F_{new} [$K_{new}/(1–R_{new})$]. For further plot details, see the caption of Figure 2.

We now turn to our main hypothesis, which was that reducing the delay (ISI; from 3500 ms to 1500 ms) between the words and math problems during the study phase would reduce the magnitude of the memory enhancement associated with emotion. As reported above, the findings of emotional memory enhancement were limited to the recollection measure and were only significant for the arousal manipulation. Since it is possible that the limited nature of these effects was due to collapsing over the delay variable, such that a weaker effect in one condition (e.g., the short delay) might have washed out a stronger effect in the other condition (e.g., the long delay), we sought to determine whether the delay manipulation interacted with either or both emotional factors. To test this, we conducted separate three-way repeated-measures ANOVAs, including both emotion factors (valence and arousal) and the delay factor (short versus long), for the recollection and familiarity measures. Because delay is of interest here, only the effects involving that factor are reported.

Figure 8. Recollection measures for the test phase, according to the valence, arousal, and delay conditions





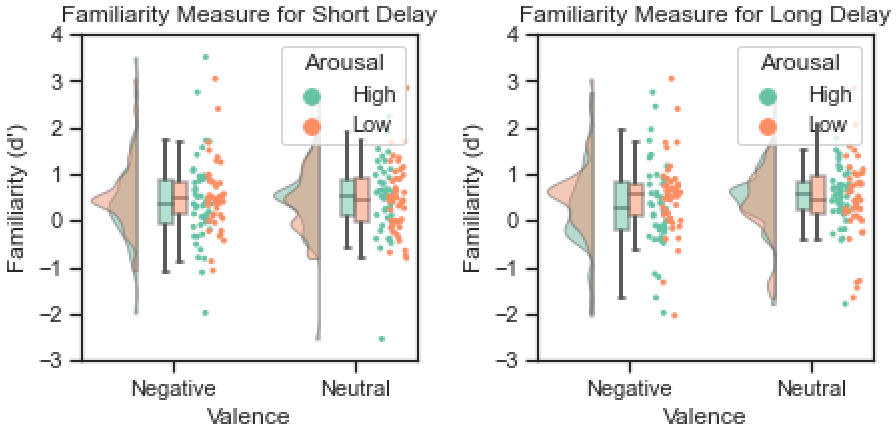
Recollection was computed as $(R_{old}–R_{new})/(1–R_{new})$ to take into consideration the overall tendency that participants had to make remember responses. The upper panel provides density plots and boxplots, along with the means for individual subjects (see caption of Figure 2 for further plot details). The lower panel provides the condition-wise means and 95% confidence interval (CI) for clarity.

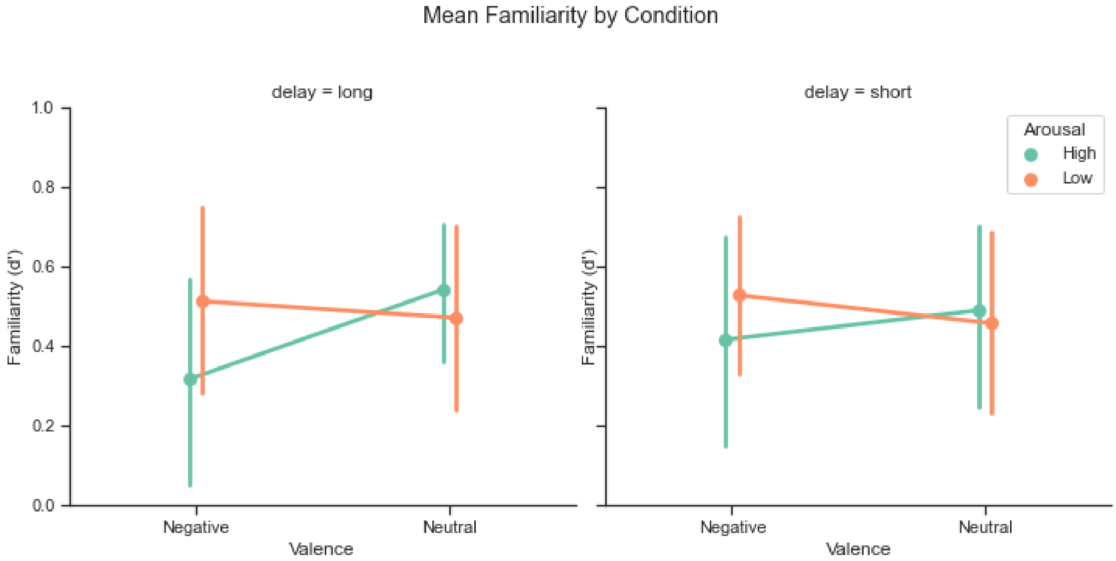
The recollection measures according to valence, arousal, and delay are summarized in Table 3 and Figure 8. As can be seen in the figures, recollection was numerically higher for the high compared to low arousal conditions, and the difference is apparent for both the long and short delay conditions, consistent with the main effect of arousal we reported earlier. Although it appears as though the arousal difference is larger for the long delay, the arousal × delay interaction did not reach significance ($F < 1$). Because the effects according to delay were central to our predictions, subsidiary ANOVAs were also conducted separately for each delay condition. These analyses revealed that the arousal effect was significant for the long delay ($F(1, 46) = 5.33$, $p = 0.026$, $\eta_p^2 = 0.10$) but not quite for the short delay ($F(1, 46) = 3.62$, $p = 0.064$, $\eta_p^2 = 0.07$).

Returning to the omnibus ANOVA, the only delay-related effect that even approached significance was the valence × delay interaction ($F(1, 46) = 3.41$, $p = 0.07$, $\eta_p^2 = 0.07$). As shown in Figure 8, however, the form of this interaction was inconsistent with our predictions, whereby the valence difference was slightly larger following the short delay. The main effect of delay and the three-way interaction for the recollection measure were also non-significant (both Fs < 1).

Turning to the familiarity measure, which is summarized with respect to each of the factors in Table 3 and Figure 9, the three-way ANOVA revealed no significant effects involving the delay factor (all Fs < 1). As shown in the figures, the data for this measure appeared to be highly variable across subjects, which likely contributed to the lack of any reliable differences.

Figure 9. Familiarity measures for the test phase, according to the valence, arousal, and delay conditions





Familiarity is reported as a d’ (difference) score based on the z-transformed values of F_{old} $[K_{old}/(1–R_{old})]$ and F_{new} $[K_{new}/(1–R_{new})]$. The upper panel provides density plots and boxplots, along with the means for individual subjects (see caption of Figure 2 for further plot details). The lower panel provides the condition-wise means and 95% CI for clarity.

Discussion

In the current study, we aimed to further understand the contribution of post-encoding persistence on the memory enhancement typically associated with emotional events and stimuli. Toward this end, participants completed an online study in which they first encoded a series of words that were either emotional in nature (negatively valenced or highly arousing) or not. Importantly, each word at encoding was followed by a simple math problem after either a long (3500 ms) or short (1500 ms) delay. Approximately 48 hours later, participants underwent a memory test for the words, in which responses were distinguished according to whether any specific details surrounding a word were recollected, or the word was judged old on the basis of familiarity (in the absence of retrieving details). Our main hypothesis was that, if persistent activation was involved in enhancing the encoding of emotional stimuli, the shorter delay between the words and math problems should limit such activation and thereby reduce the enhancement. Moreover, based on several previous studies (e.g., Kensinger & Corkin, 2003, 2004; Mickley & Kensinger, 2008; for review, see Bowen et al., 2018), the enhancement should be evident primarily for recollection- as opposed to familiarity-based memory.

Partially supporting our hypotheses, we found limited evidence in support of the emotional manipulations (valence and arousal) affecting memory. In particular, the only significant effect we observed was due to the arousal manipulation on the measure of recollection, in which participants reported greater frequency of recollection for words from the high- compared to low-arousal conditions. Although this effect held up when considering the data from the long-delay condition on its own, the small size of the effect likely played a role in failing to observe an interaction between arousal and delay. Moreover, throughout all our analyses, emotional valence gave rise to no significant differences (cf. Bowen et al., 2018). These findings are thus inconsistent with those of several previous studies that have exhibited strong emotional enhancements on episodic memory, such that negatively-valenced stimuli are retrieved at higher rates than neutral stimuli, and stimuli eliciting higher ratings of arousal are also associated with better memory performance (e.g., Bradley et al., 1992; Christianson, 1992; Libkuman et al., 2004; Migita et al., 2011; Samide et al., 2019). Our ability to detect any decreases in emotional effects due to the delay manipulation, therefore, was likely limited by not establishing these typically-strong enhancements.

Given that the current study provided limited evidence of an emotional memory enhancement, particularly with respect to valence, the discussion here focuses on several potential explanations of the overall absence of strong effects. The first, and perhaps most obvious factor is the potential lack of statistical power. To that issue, our study was designed with several other studies of emotional memory in mind. Most relevant was a study by Kensinger and Corkin (2003) that both used word stimuli and employed a remember/know procedure during retrieval to distinguish recollection from familiarity. Across six experiments in that study, the number of participants ranged from 16 to 20. Here, we used a target sample size (of N ≈ 50) that was more than twice as large, along with stimulus sets per condition that were comparable (25-70 stimuli in the Kensinger & Corkin study versus our sets of 40 stimuli per condition). Together, these design features presumably gave us the statistical power necessary to detect the effects of emotion. Notably, though, it is possible that including a secondary math task decreased the overall size of any emotional enhancement. That is, even in the long delay condition, the total time of about 4 seconds to study each word (750 ms word presentation + 3500 ms ISI) may have been adversely affected by the inclusion of math problems. However, as there was no other existing study that used such a manipulation, it was difficult to plan accordingly for adequate statistical power.

Another potential limitation to our ability to detect the emotional memory enhancement is due to the use of word stimuli, which were drawn from the database of

Warriner et al. (2013). As noted in the Introduction, most of the existing research on emotional memory has employed pictorial stimuli that often come from the IAPS norms (Lang et al., 2008). Our decision to instead use words was motivated by multiple factors. First, we wanted to establish emotional effects with stimuli of limited display size and visual differences (e.g., intensity, color) in order to eventually use the behavioral procedures alongside the acquisition of electroencephalographic (EEG) data. Although pictorial stimuli might produce stronger memory effects, the eye movements associated with viewing pictures could pose problems when assessing for EEG correlates. Secondly, while Bradley and Lang (1999) obtained valence and arousal ratings for word stimuli as part of the Affective Norms for English Words (ANEW) database, and those stimuli have been previously used in several studies (e.g., Kensinger & Corkin, 2003), we wanted to incorporate updated norms that could have changed in the decades since. In particular, Warriner et al. (2013) provide standard valence, arousal, and dominance norms in addition to data that allowed us to control for gender differences in ratings – a factor that could have been particularly relevant given our predominantly female participant sample.

Ultimately, the choice to use word stimuli could have limited our ability to detect differences in emotional memory, which are likely to be stronger with the salient picture stimuli that have traditionally been used. Additionally, our stimulus selection procedures from the larger database might have not been ideal in some respect, such as failing to control for conceptual similarity among multiple stimuli. Any such differences between conditions could have, in turn, affected participants' memories and the corresponding recollection and familiarity measures derived from their responses (Montefinese et al., 2015; Tompary & Thompson-Schill, 2020). Despite these overlooked and potentially subtle differences, it is also possible that the valence and arousal ratings of words are not capturing the same stimulus qualities or degrees of those qualities as the ratings with pictures do. Indeed, many of the studies demonstrating strong emotional memory advantages for pictures have confounded valence and arousal, such that negative stimuli are often high in arousal and neutral stimuli are low in arousal, further obscuring whether the effects load on a particular dimension (e.g., Harris & Pashler, 2005; Heuer & Reisberg, 1990; Migita et al., 2011; for such a confound with words, see Kensinger & Corkin, 2003; Sharot & Phelps, 2004). Nonetheless, it is useful to further explore the emotional enhancement of memory, as we have done here, particularly if the phenomenon is limited to specific types of experimental materials.

A final possible explanation of the limited findings reported here involves general concerns with control during data collection. Specifically, as this study was conducted online, we had to remove several participants for failing to complete both portions of the experiment within the allotted time period (i.e., within a 36- to 60-hour window). Since participants also completed this study in the location of their choosing, we had no control over environmental distractions, which could have led to effects of divided attention (Naveh-Benjamin et al., 2003; Greene et al., 2021). Additionally, due to current events such as the COVID-19 pandemic, and social and civil unrest, general emotional well-being could have contributed in undesirable ways to the effects of interest here. That is, many of the emotional words we expected to produce strong recollection outcomes – such as “deathly”, “infectious”, and “violent” – were likely encountered frequently over the course of data collection, potentially restricting participants’ ability to distinguish the sources of these stimuli (Otani et al., 2013; Perrson et al., 2016). Similarly, the time of day participants completed the experiment could be a factor, such that people tend to be more aroused in the morning than in the evening which could affect encoding and later retrieval of information (e.g., Baddeley et al., 1970).

To summarize, contrary to our hypothesis and the NEVER model proposed by Bowen et al. (2018), interrupting post-stimulus encoding via a secondary task had no effect on the memory enhancement associated with negative valence. To our knowledge, this was the first experiment designed to directly test that model. Although we failed to support it, the findings of the current study were somewhat consistent with the memory advantage that has been previously shown to be due to high arousal (e.g., Bradley et al., 1992; Christianson, 1992). As we have discussed here, several potential limitations of the current study should be addressed in future research, including those related to the experimental control afforded by a laboratory setting and consideration of factors such as participants’ general mood and well-being (e.g., sleep patterns), which might have affected emotional processing and memory performance.

References

Addante, R. J., de Chastelaine, M., & Rugg, M. D. (2015). Pre-stimulus neural activity predicts successful encoding of inter-item associations. *Neuroimage, 105*, 21-31. https://doi.org/10.1016/j.neuroimage.2014.10.046

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America, 91*(15), 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Baddeley, A. D., Hatter, J. E., Scott, D., & Snashall, A. (1970). Memory and time of day. *The Quarterly Journal of Experimental Psychology*, *22*(4), 605-609.

Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *The Journal of Neuroscience, 31*(24), 9032-9042. https://doi.org/10.1523/JNEUROSCI.0702-11.2011

Ben-Yakov, A., Eshel, N., & Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *Journal of Experimental Psychology: General, 142*(4), 1255-1263. https://doi.org/10.1037/a0033558

Ben-Yakov, A., Dudai, Y., & Mayford, M. R. (2015). Memory retrieval in mice and men. *Cold Spring Harbor Perspectives in Biology, 7*(12), 1-29. https://doi.org10.1101/cshperspect.a021790

Bowen, H. J., Kark, S. M., & Kensinger, E. A. (2018). NEVER forget: Negative emotional valence enhances recapitulation. *Psychonomic Bulletin & Review, 25*, 870-891. https://doi.org/10.3758/s13423-017-1313-9

Bowen, H. J., & Kensinger, E. A. (2017). Recapitulation of emotional source context during memory retrieval. *Cortex, 91*, 142-156. https://doi.org/10.1016/j.cortex.2016.11.004

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 379–390. https://doi.org/10.1037/0278-7393.18.2.379

Bradley, M.M., & Lang, P.J. (1999). Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.

Brainerd, C. J. (2018). The emotional-ambiguity hypothesis: A large-scale test. *Psychological Science, 29*(10), 1706-1715. https://doi.org/10.1177/0956797618780353

Brainerd, C. J., & Bookbinder, S. H. (2019). The semantics of emotion in false memory. *Emotion, 19*(1), 146-159. https://doi.org/10.1037/emo0000431

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281*(5380), 1185–1187. 10.1126/science.281.5380.1185

Christianson, S.-Å. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112*(2), 284–309. https://doi.org/10.1037/0033-2909.112.2.284

Cohen, N., Pell, L., Edelson, M. G., Ben-Yakov, A., Pine, A., & Dudai, Y. (2015). Peri-encoding predictors of memory encoding and consolidation. *Neuroscience & Biobehavioral Reviews, 50*, 128-142. https://doi.org/10.1016/j.neubiorev.2014.11.002

Cohn, M., Moscovitch, M., Lahat, A., & McAndrews, M. P. (2009). Recollection versus strength as the primary determinant of hippocampal engagement at retrieval. *Proceedings of the National Academy of Sciences, 106*(52), 22451-22455. https://doi.org/10.1073/pnas.0908651106

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin, 136*(1), 87–102. https://doi.org/10.1037/a0017937

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion, 1*(1), 5-11. https://doi.org/10.1037/1528-3542.1.1.5

Ford, J. H., Morris, J. A., & Kensinger, E. A. (2014). Effects of emotional valence on the neural correlates of episodic memory and elaboration. *Journal of Cognitive Neuroscience, 26*(4), 825-839. https://doi.org/10.1162/jocn_a_00529

Ford, J. H., Verfaellie, M., & Giovanello, K. S. (2010). Neural correlates of familiarity-based associative retrieval. *Neuropsychologia, 48*(10), 3019-3025. https://doi.org/10.1016/j.neuropsychologia.2010.06.010

Greene, N. R., Martin, B. A., & Naveh-Benjamin, M. (2021). The effects of divided attention at encoding and at retrieval on multidimensional source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(11), 1870–1887. https://doi.org/10.1037/xlm0001051

Hulbert, J. C., Henson, R. N., & Anderson, M. C. (2016). Inducing amnesia through systemic suppression. *Nature Communications, 7*(11003), 1-9. https://doi.org/10.1038/ncomms11003

Hulbert, J. C., Hirschstein, Z., Brontë, C. A. L., & Broughton, E. (2018). Unintended side effects of a spotless mind: Theory and practice. *Memory, 26*(3), 306-320. https://doi.org/10.1080/09658211.2017.1354999

Hulse, L. M., Allan, K., Memon, A., & Read, J. D. (2007). Emotional arousal and memory: A test of the poststimulus processing hypothesis. *The American Journal of Psychology, 20*(1), 73-90. https://doi.org/10.2307/20445382

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex 17*(11), 2507–2515. https://doi.org/10.1093/cercor/bhl156

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. *The Journal of Neuroscience, 24*(17), 4172–4180. https://doi.org/10.1523/JNEUROSCI.0624-04.2004

Kensinger, E. A., Addis, D. R., & Atapattu, R. K. (2011). Amygdala activity at encoding corresponds with memory vividness and with memory for select episodic details. *Neuropsychologia, 49*, 663-673. https://doi.org/10.1016/j.neuropsychologia.2011.01.017

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*(8), 1169-1180. https://doi.org/10.3758/BF03195800

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences, 101*(9), 3310-3315. https://doi.org/10.1073/pnas.0306408101

Kensinger, E. A., & Ford, J. H. (2020). Retrieval of emotional events from memory. *Annual Review of Psychology, 71*, 251-272. https://doi.org/10.1146/annurev-psych-010419-051123

Kensinger, E. A., & Schacter, D. L. (2008). Neural processes supporting young and older adults' emotional memories. *Journal of Cognitive Neuroscience, 20*(7), 1161–1173. https://doi.org/10.1162/jocn.2008.20080

Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology, 65*(2), 190–193. https://doi.org/10.1037/h0040288

Kluyver, T., Ragan-Kelley, B., Fernando Perez, Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdala, S., & Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90). 10.3233/978-1-61499-649-1-87

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*, 261-273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Leiker, E. K., & Johnson, J. D. (2014). Neural reinstatement and the amount of information recollected. *Brain Research, 1582*, 125-138. https://doi.org/10.1016/j.brainres.2014.07.026

Libkuman, T. M., Stabler, C. L., & Otani, H. (2004). Arousal, valence, and memory for detail. *Memory (Hove, England)*, *12*(2), 237–247. https://doi.org/10.1080/09658210244000630

Madan, C. R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability? *Psychonomic Bulletin & Review, 28*, 583-595. https://doi.org/10.3758/s13423-020-01820-w

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Mickley Steinmetz, K. R., & Kensinger, E. A. (2009). The effects of valence and arousal on the neural activity leading to subsequent memory. *Psychophysiology, 46*(6), 1190–1199. https://doi.org/10.1111/j.1469-8986.2009.00868.x

Migita, M., Otani, H., Libkuman, T. M., & Sheffert, S. M. (2011). Preattentive processing, poststimulus elaboration, and memory for emotionally arousing stimuli. *The Journal of General Psychology*, *138*(4), 260–280. https://doi.org/10.1080/00221309.2011.604364

Montefinese, M., Zannino, G. D., & Ambrosini, E. (2015). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological Research*, *79*(5), 785–794. https://doi.org/10.1007/s00426-014-0615-z

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Naveh-Benjamin, M., Craik F. I. M., Perretta J. G., & Tonev S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *The Quarterly Journal of Experimental Psychology Section A*. *53*(3), 609-625. https://doi.org/10.1080/713755914

Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition, 31*(7), 1021-1035. https://doi.org/10.3758/BF03196123

Naveh-Benjamin, M., & Kilb, A. (2012). How the measurement of memory processes can affect memory performance: The case of remember/know judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 194–203. https://doi.org/10.1037/a0025256

Otani, H., Libkuman, T. M., Goernert, P. N., Kato, K., Migita, M., Freehafer, S. E., & Landow, M. P. (2012). Emotion, directed forgetting, and source memory. *British Journal of Psychology (London, England : 1953)*, *103*(3), 343–358. https://doi.org/10.1111/j.2044-8295.2011.02078.x

Paller K. A., & Wagner A. D. (2002) Observing the transformation of experience into memory. *Trends in Cognitive Science, 6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. https://doi.org/10.3758/s13428-018-01193-y

Persson, B. M., Ainge, J. A., & O'Connor, A. R. (2016). Disambiguating past events: Accurate source memory for time and context depends on different retrieval processes. *Neurobiology of learning and memory*, *132*, 40–48. https://doi.org/10.1016/j.nlm.2016.05.002

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89-102. https://doi.org/10.3758/BF03211168

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic-episodic distinction. *Trends in Cognitive Sciences, 23*(12), 1041-1057. https://doi.org/10.1016/j.tics.2019.09.008

Rolls, E. T. (2000). Hippocampo-cortical and cortico-cortical backprojections. *Hippocampus, 10*(4), 380–388. https://doi.org/10.1002/1098-1063(2000)10:4<380::AID-HIPO4>3.0.CO;2-0

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience, 12*(4), 664-678. https://doi.org/10.1162/089892900562291

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. *Progress in Brain Research, 169*, 339–352. https://doi.org/10.1016/S0079-6123(07)00021-0

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia, 50*(13), 3070-3079. https://doi.org/10.1016/j.neuropsychologia.2012.06.004

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality & Social Psychology*, *39,* 1161-1178. https://doi.org/10.1017/S0954579405050340

Samide, R., Cooper, R. A., & Ritchey, M. (2019). A database of news videos for investigating the dynamics of emotion and memory. *Behavior Research Methods, 52*, 1469-1479. https://doi.org/10.3758/s13428-019-01327-w

Sharot, T., & Phelps, E. A. (2004) How arousal modulates memory: disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 294–306. https://doi.org/10.3758/CABN.4.3.294

Sharot, T., & Yonelinas, A. P. (2008) Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition 106*(1), 538–547. https://doi.org/10.1016/j.cognition.2007.03.002

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Tompary, A., & Thompson-Schill, S. L. (2020, February 4). Semantic influences on episodic memory distortions. https://doi.org/10.31234/osf.io/u754s

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne, 26*(1), 1–12. https://doi.org/10.1037/h0080017

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071

Vaish, A., Grossman, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383-403. https://doi.org/10.1037/0033-2909.134.3.383

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vilberg, K. L., & Rugg, M. D. (2009). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping, 30*(5), 1490-1501. https://doi.org/10.1002/hbm.20618

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281*(5380), 1188–1191. https://doi.org/10.1126/science.281.5380.1188

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Yonelinas, A. P., Kroll,N. E., Dobbins, I., Lazzara,M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of “remember”–“know,” process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*, 323–339. https://doi.org/10.1037//0894-4105.12.3.323

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*(3), 361-379. https://doi.org/10.1037/0096-3445.130.3.361

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yu, S. S., Johnson, J. D., & Rugg, M. D. (2012). Hippocampal activity during recognition memory co-varies with the accuracy and confidence of source memory judgements. *Hippocampus, 22*(6), 1429-1437. https://doi.org/10.1002/hipo.20982

Zellner, A., Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In:Bernardo, J. M., Degroot, M. H., Lindley, D. V., Smith, A. F. M. (Eds.), *Bayesian Statistics: Proceedings of the first international meeting held in Valencia*. University of Valencia, pp. 585–603. https://doi.org/10.1007/BF02888369