Semantic Relatedness or Arousal: What Drives Enhancement of Memory for Emotional Words?

by

Brian Sungsoon Kim

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Emotional stimuli are typically better remembered than neutral ones. Two accounts have been proposed for the advantage in memory for lists of words. The first is that it arises due to enhanced arousal, and the second is that it arises due to the inherent semantic relatedness of emotional words. We compared these accounts by examining memory performance for lists of words in which arousal and relatedness were manipulated. In three experiments (n = 418), conducted online, participants were asked to encode lists containing words that were either unrelated, related, or emotionally valenced, followed by free recall tests. List type was manipulated within-subject. Participants also rated the arousal and emotional valence they experienced for each word using a Self-Assessment Manikin (SAM). Latent semantic analysis (in Experiment 1) and free association norms (in Experiment 2) were used to match the semantic relatedness of words in the emotional (negative) and related lists. Free recall of related and negatively valenced word lists was significantly higher than for unrelated lists. In Experiment 3, we used free association norms to create emotional lists of mixed valence (negative and positive), effectively lowering semantic relatedness within the list, but keeping arousal high. Despite lowering relatedness of the emotional word lists, recall of related and emotional word lists was again of similar magnitude, with memory for both significantly higher than for unrelated lists. Importantly, using logistic regression we showed that participants' arousal ratings predicted recall in all three experiments. Findings suggest that arousal, not semantic relatedness, explains why memory is better for lists of emotional words.

Acknowledgements

First and foremost, I'd like to thank the two best supervisors one could ask for, Dr. Myra Fernandes and Dr. Katherine White. I consider myself truly fortunate to have had the opportunity to work with both of them over the past couple of years. Reflecting on this time, I am deeply grateful for their unwavering support, invaluable guidance, and constant encouragement. Their mentorship has profoundly shaped not only my research but also the way I think and work, and I will be able to carry their teachings with me throughout my career. It was truly an honour and an absolute pleasure to work with both of you.

I would also like to thank everyone I met throughout my graduate student journey. In particular, Khalil Husein, Sophia Tran, Yadu Sivashankar, Patrick Tsapoitis, Jisoo Kim, and Miranda Chan – thank you for your camaraderie, your support, the laughs, and the many good times we shared helping one another navigate the days of grad school in PAS.

To my partner and best friend, Clara Li – thank you for your endless love, patience, and encouragement. Your belief in me keeps me grounded and motivated. I'm so grateful to have you by my side.

Finally, I want to express my heartfelt thanks to my family. To my sisters, Jeannie and Adrienne Kim, thank you for always being there to cheer me on and for reminding me of the bigger picture. To my mom, Insun Kim, thank you for your boundless warmth, selflessness, and strength – your wisdom continues to inspire me. And to my dad, Sunchul Kim, thank you for your quiet perseverance, steady support, and the values you've instilled in me that guide me every day – your care and love has been the foundation of everything I do. 엄마, 아빠, 감사합니다. 사랑합니다.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Emotional stimuli are typically better remembered than neutral ones, an effect referred to as Emotional Enhancement of Memory (EEM; Finkenauer et al., 1998; Hamann, 2001; Hamann et al., 1999; Rapaport, 1943; Rubin & Friendly, 1986). Broadly, affective stimuli can be described as varying along two dimensions: valence (negative to positive) and arousal (low to high). Emotional stimuli, characterized by high arousal, are more accurately, vividly and quickly recalled than neutral stimuli. They are also retained in memory for longer periods of time (Cahill & McGaugh, 1995; Kensinger, 2004; Mather, 2007). The EEM has been demonstrated for different types of stimuli, such as pictures (Blake et al., 2001), movies and scenes (Subramanian et al., 2014), music (Aubé et al., 2013) and word lists (Doerksen & Shimamura, 2001; Kensinger & Corkin, 2003; Majerus & D'Argembeau, 2011). This thesis focuses on the EEM for lists of words.

1.1 Role of Arousal

Identifying the mechanism underlying the EEM effect has garnered significant interest. Converging evidence from studies involving neurologically intact humans, patients, and animals highlights the critical role of certain brain structures, particularly the amygdala, in the processing and storage of emotional memories (Bechara et al., 1995; Kapp et al., 1979; LeDoux, 1994). Physiologically, emotional stimuli have been shown to elicit higher levels of arousal (Eysenck, 1976; LaBar & Phelps, 1998). Specifically, Cahill and McGaugh (1998) emphasize the role of stress hormones, specifically adrenaline and cortisol, released in response to emotional stimuli. These hormones serve two key adaptive functions: (1) facilitating an immediate response to the arousing item or event and (2) promoting the long-term consolidation of memories after exposure to emotional stimuli. The immediate response involves the activation of the sympathetic nervous system, preparing the organism for "fight or flight". This short-term function aids in survival by enabling rapid physiological responses to the emotional stimulus (Ozawa et al., 2022).

Beyond the immediate response, these stress hormones also play a key role in long-term memory consolidation. Specifically, Cahill and McGaugh (1998) describe how stress hormones modulate the hippocampus, prefrontal cortex and amygdaloid complex, thus improving the likelihood that emotionally significant events will be remembered. One explanation for this bias stem from evolutionary pressures - humans may prioritize attending to negative stimuli in particular as this could be critical for survival (Brown & Kulik, 1977; Talarico et al., 2009), such as remembering which berries lead to food poisoning. This account has been extended to explain the superior recall of emotional sets of words compared to neutral sets.

1.2 Limitations of the Memory Consolidation Account for Immediate EEM

There is substantial evidence that arousal levels increase following exposure to emotional stimuli. However, one crucial limitation in the arousal-based consolidation explanation of the EEM effect is that the processes involved (e.g. action of elevated levels of stress hormones in the brain) take an appreciable time to occur. There is behavioural evidence to suggest that when arousal is manipulated at encoding and recall is tested at multiple time intervals (ranging from an hour to days), the memory benefit for highly arousing stimuli is greater after longer time delays (Eysenck, 1976; Burke et al., 1992; Hamann et al., 1999). However, this long-term consolidation mechanism cannot readily explain the robust finding of an immediate EEM reported by many other investigators (Blake et al., 2001; Bradley et al., 1992; Hamann et al., 1999; Kensinger et al., 2002; LaBar & Phelps, 1998; MacKay et al., 2004). Furthermore, it appears that the amygdala is not always required for EEM. Research demonstrates that amygdala-lesioned patients still exhibit immediate EEM (LaBar & Phelps, 1998) suggesting the possible existence of an alternative or dual route to EEM, not explained by the consolidation mechanism of arousal alone. Hamann (2001) also highlights the conflict between long-acting norepinephrine effects (lasting several minutes at least and remaining elevated during presentation of neutral stimuli in mixed list contexts) and the selective enhancement of emotional stimuli. If elevated norepinephrine alone were sufficient to improve memory, then one would expect the benefit to extend to neutral stimuli presented in the context of emotional stimuli.

In light of these limitations in the consolidation account of enhanced memory for emotional words, researchers have proposed an alternative mechanism to help explain the immediate EEM benefit. Critically, the enhanced immediate recollection of emotional items has been shown to correlate with amygdala and noradrenergic system activity during encoding (Waring & Kensinger, 2011). This mechanism focuses on the cognitive factors inherent to emotional stimuli that can mediate memory such as increased attention. Ample evidence has shown that emotional stimuli attract increased attention compared to neutral stimuli during encoding (Fox et al., 2001; MacKay et al., 2004; Schimmack & Derryberry, 2005; Williams et al., 1996). The increased activation in anterior temporal and extrastriate visual cortex while viewing emotional pictures is attributed to increased attentional and perceptual processing of the emotional material (e.g. Bradley et al., 1992; Mourão-Miranda et al., 2003). Because allocation of attention during encoding is known to be an important determinant of memory (Craik et al., 1996; Smith, 1895), emotional stimuli may be remembered better because they attract more attention during the encoding phase (Hamann, 2001; Kensinger & Corkin, 2003). Thus, the immediate superior recognition of emotional items could be attributed primarily to their attraction of attention during encoding which is also thought to be based on the amygdala and the central noradrenergic system.

1.3 Role of Semantic Relatedness

Another factor known to enhance memory is semantic relatedness. Research reliably illustrates that semantically related stimuli are better remembered than unrelated ones (Mandler, 1967; Poirier & Saint-Aubin, 1995; Tulving & Pearlstone, 1966). This phenomenon can be explained by the Fuzzy-Trace Theory (FTT; Brainerd & Reyna, 1998), which distinguishes between two types of memory traces: gist and verbatim. A gist-based memory system operates by extracting the overarching theme or general meaning after exposure to multiple exemplars of the same category (Poirier & Saint-Aubin, 1995). This process reduces cognitive load as only the key theme needs to be retained in memory, ultimately enhancing recall of all items related to the gist (Bookbinder & Brainerd, 2017). This is the way in which semantically related stimuli are better remembered than unrelated stimuli. In contrast, a verbatim memory system stores each item individually, making it more sensitive to the availability of cognitive resources (Bookbinder & Brainerd, 2017). Importantly, the distinction between gist and verbatim memory systems has consequences depending on the type of test used to assess memory. For instance, reliance on gist memory with categorized word lists typically benefits performance on free recall tests, as the overarching theme or general meaning can be used as a memory cue during retrieval. On the other hand, gist-based memory typically impairs performance on recognition tests, possibly due to interference from similar items (Koutstaal & Schacter, 1997; Roediger & McDermott, 1995). In the present study we assessed memory for list of words using a free recall test, which should benefit lists that are semantically interrelated (Barnhardt et al., 2006).

Talmi and Moscovitch (2004) proposed that immediate memory for emotional words might also be heightened by the inherent semantic relatedness of emotional words when presented in a list. They argued that many studies comparing memory for emotional and neutral words often overlook a critical confound: emotional words are typically more semantically related than randomly selected neutral words. This can skew memory performance. Indeed, research has shown that lists of negative emotional words are more semantically related than lists of neutral words (Kazanas & Altarriba, 2015; Kuperman et al., 2014). Similar concerns were raised by Dougal and Rotello (2007), who argued that negative emotional word lists could form a cohesive memory trace with a shared emotional tone, unlike neutral word lists which lack such a unifying feature. This shared "gist" for emotional words provides an additional recall advantage not available to neutral words. To address this potential confound, Talmi and Moscovitch (2004) conducted a series of experiments comparing memory performance across three types of word lists: negatively valenced emotional words, unrelated neutral words and, novelly, related neutral words. Their study revealed that the memory benefit for semantically related and emotional word lists, relative to unrelated lists, was similar in magnitude. Based on their results they suggested that the memory advantage typically observed for emotional words might be partially attributable to their semantic relatedness rather than arousal-level alone.

Semantic relatedness has also been implicated as a contributing factor to the EEM effect during retrieval processes (Polyn et al., 2009). One way this influence manifests is through the tendency for semantically related items to be recalled together, a phenomenon known as semantic clustering. Although stimuli are typically encoded in a randomized order, participants often retrieve items from the same semantic category in succession (e.g., recalling clothing items before utensil items). This tendency can be quantified using the List-Based Semantic Clustering Index (LBCsem; Stricker et al., 2002). Prior studies have shown that such clustering enhances memory performance (Tulving, 1962). Interestingly, this effect extends to emotional stimuli. Talmi et al. (2007), for instance, found that participants who exhibited stronger semantic clustering during free recall showed no memory advantage for emotional over neutral stimuli, whereas those who did not cluster showed a typical EEM effect. This suggests that semantic organization at retrieval may level the playing field by boosting recall for neutral content. Taken together, these findings indicate that mechanisms beyond arousal, such as semantic relatedness, may play a critical role in emotional memory enhancement.

In the present study, we aimed to determine whether the robust immediate memory advantage observed for emotional word lists is driven primarily by an increased arousal state, which has long been considered the primary mechanism, or whether it is instead attributable to the inherent semantic relatedness of emotional words, as suggested by Talmi and Moscovitch (2004). To address this question, we conducted three experiments. In Experiment 1, we aimed to replicate the findings of Talmi and Moscovitch (2004) while extending their work to a new set of word lists not used in their original study. In Experiment 2, we opted to examine a different operational definition of semantic relatedness to determine whether this changed the pattern of results. Specifically, we assessed relatedness using free association norms rather than latent semantic analysis. The goal was to operationalize semantic relatedness in different ways to assess the replicability and generalizability of any effect. Finally, in

Experiment 3 we introduced a novel manipulation to disentangle arousal and semantic relatedness: we created an emotional word list with mixed valence (an equal proportion of positive and negative words). This approach effectively reduced semantic relatedness while maintaining high arousal, providing a direct examination of the contribution of relatedness to the EEM effect.

Chapter 2
Experiment 1

The goals of Experiment 1 were to replicate the findings of Talmi and Moscovitch (2004) as well as to examine whether the findings generalize to another set of word lists. To this end, we used two sets of stimuli. One set (referred to henceforth as Experiment 1a) used the same 3 lists of words (unrelated, related and emotional) as in Experiment 3 of the Talmi and Moscovitch (2004) study. A second set (referred to henceforth as Experiment 1b), was created to assess the generalizability and replicability of their findings. To this end, we created 3 new word lists, again consisting of unrelated, related and emotional words, selected from the Medical Research Council (MRC; Wilson, 1988) psycholinguistic database. These two sets of words were matched for various lexical characteristics, with the exception that lists in Experiment 1b were somewhat lower in frequency.

For all lists in Experiment 1, we measured semantic relatedness using latent semantic analysis (LSA; Landauer et al., 1998, accessed at http://lsa.colorado.edu). The LSA modeling technique is built on a corpus containing 104,852 terms and 942,425 documents. It uses a mathematical function to analyze word co-occurrence within this corpus, representing each word as a vector in a multi-dimensional space. LSA identifies topics and categories within the text corpus and assigns a similar measure to pairs of words based on their contextual alignment. Similarity is calculated as the cosine of the angle between the vectors of two words, ranging from -1 (maximum dissimilarity) to 1 (maximum similarity). This measure provides a robust, quantitative estimate of semantic relatedness across a wide range of contexts (Landauer et al., 1998).

2.1 Methods

2.1.1 Participants

We collected data from 141 participants using the online data collection platform Prolific (prolific.com). This sample size was chosen to ensure sufficient power to detect a medium-sized effect in the analysis at an alpha level of .05 based on prior research. One participant was excluded following a compliance check, bringing the total number of participants to 140. Experiment 1a consisted of 71 participants (34 males, 37 females). Ages ranged from 22 to 71 years (*M* = 41.26, *SD* = 11.98). Experiment 1b consisted of 69 participants (28 males, 41 females). Ages ranged from 21 to 76 years (*M* = 42.17, *SD* = 12.59). Participants received remuneration at a rate of £6 per hour. All participants reported their highest level of education completed as an undergraduate degree. All participants reported being monolingual in English and raised exclusively in English.

2.1.2 Materials

In Experiment 1a, stimuli consisted of the same three 16-word lists used in Experiment 3 of Talmi and Moscovitch (2004); one list contained unrelated words, one contained semantically related words, and one contained negatively valenced emotional words. For Experiment 1b, we created another three 16-word lists (unrelated, semantically related and emotional) using the MRC psycholinguistic database (Wilson, 1988). We controlled lexical variables such as concreteness, familiarity, imageability and number of letters across experiments, as they have been known to influence memory (Balota & Neely, 1980). Arousal and valence values were determined based on the English Lexicon Project (Balota et al., 2007). See Table 1 for means and standard deviations of lexical variables.

Within Experiment 1a, one-way ANOVAs confirmed that word lists were matched for the following lexical variables (see Table 1 for means): concreteness: $F(2, 45) = 1.64$, $p = .21$, $\eta_p^2 = 0.02$; familiarity: $F(2, 45) = 1.94$, $p = .16$, $\eta_p^2 = 0.08$; imageability: $F(2, 45) = 1.63$ $p = .21$, $\eta_p^2 = 0.02$; frequency: $F(2, 45) = 0.75$, $p = .48$, $\eta_p^2 = 0.03$; and number of letters: $F(2, 45) = 0.66$, $p = .52$, $\eta_p^2 = 0.03$. Importantly, however, there was a significant difference in valence across word lists, $F(2, 45) = 99.26$, $MSE = 0.52$, $p < .001$, $\eta_p^2 = 0.82$ with emotional word lists significantly lower in valence (more negative) than the unrelated ($p < .001$) and related list ($p < .001$). There was also a significant difference in arousal across word lists, $F(2, 45) = 62.49$, $MSE = 0.48$, $p < .001$, $\eta_p^2 = 0.74$ with emotional words significantly higher in arousal than the unrelated ($p < .001$) and related ($p < .001$) word lists. Finally, the word lists differed significantly in relatedness, $F(2, 47) = 40.30$, $MSE = 0.22$, $p < .01$, $\eta_p^2 = 0.66$ such that the LSA scores for the unrelated word list ($M_{LSA} = 0.10$) were significantly lower than for the related ($M_{LSA} = 0.25$), $p < .01$, and emotional ($M_{LSA} = 0.21$), $p < .01$ lists.

Similarly, within Experiment 1b, word lists were matched for the following lexical variables: concreteness: $F(2, 45) = 2.64$, $p = .08$, $\eta_p^2 = 0.10$; familiarity: $F(2, 45) = 1.63$, $p = .21$, $\eta_p^2 = 0.07$; imageability: $F(2, 45) = 0.83$, $p = .44$, $\eta_p^2 = 0.04$; frequency: $F(2, 45) = 0.65$, $p = .53$, $\eta_p^2 = 0.03$; and number of letters: $F(2, 45) = 0.67$, $p = .51$, $\eta_p^2 = 0.03$. Importantly, a one-way ANOVA revealed a significant difference in valence across word lists, $F(2, 45) = 146.77$, $MSE = 0.37$, $p < .001$, $\eta_p^2 = 0.87$, with emotional word lists significantly lower in valence than the unrelated ($p < .001$) and related list ($p < .001$). There was also a significant difference in arousal across word lists, $F(2, 45) = 58.96$, $MSE = 0.42$, $p < .001$, $\eta_p^2 = 0.72$, with emotional words significantly higher in arousal than the unrelated ($p < .001$) and related ($p < .001$) word lists. Finally, the word lists differed in relatedness, $F(2, 47) = 23.75$, $MSE = 0.74$, $p < .01$, $\eta_p^2 = 0.62$ such that the LSA scores for the unrelated word list ($M_{LSA} = 0.08$) were significantly lower than for the related ($M_{LSA} = 0.22$), $p < .01$, and emotional ($M_{LSA} = 0.24$), $p < .01$ lists.

Additionally, collapsed across word list type, words across Experiments 1a and 1b were matched on several variables. Independent *t* tests indicated no significant differences between experiments in terms of familiarity $t(94) = -0.07$, $SE = 9.72$, $p = .95$, imageability $t(94) = 1.44$, $SE = 5.83$, $p = .15$, number of letters $t(94) = 0.88$, $SE = 0.26$, $p = .38$, valence $t(94) = 0.78$, $SE = 0.23$, $p = .44$, arousal $t(94) = 1.65$, $SE = 1.23$, $p = .10$ and LSA, $t(94) = 0.98$, $SE = 1.52$, $p = .33$. However, concreteness, $t(94) = 3.94$, $SE = 12.42$, $p < .001$ and frequency $t(94) = 4.54$, $SE = 3.45$, $p < .001$ did significantly differ between experiments, in that Experiment 1b had lower frequency words, which also tend to be less concrete.

Table 1

Lexical variables for word lists in Experiment 1

	Experiment 1a			Experiment 1b		
Lexical variables	Unrelated	Related	Emotional	Unrelated	Related	Emotional
Concreteness	512.81 (68.1)	541.63 (65.49)	527.19 (94.59)	569.19(48.69)	590.13 (31.14)	535.25 (81.51)
Familiarity	520.75 (51.17)	550.19 (48.42)	543.63 (29.75)	556.44 (45.63)	533.94 (44.90)	526.19 (38.04)
Imageability	545 (77.07)	551.75 (55.59)	548.56 (47.41)	553.38 (47.40)	575.13 (35.37)	570.06 (41.23)
Frequency	53.31 (30.93)	56.44 (29.14)	67.38 (39.90)	32.38 (10.90)	36.44 (15.47)	35 (17.13)
NLET	5.13 (1.03)	5.06 (1.47)	4.75 (1.09)	5.00 (1.11)	4.81 (0.49)	2.56 (0.50)
Valence	5.51 (0.49)	5.41 (1.00)	2.34 (0.47)	5.54 (0.59)	5.95 (0.49)	2.56 (0.5)
Arousal	3.43 (0.88)	4.14 (0.58)	6.07 (0.55)	3.49 (0.37)	3.32 (0.55)	5.56 (0.94)
LSA	0.10	0.25	0.21	0.08	0.22	0.24

Following the memory task, participants completed the LexTALE as a measure of their English proficiency (Lemhöfer & Broersma, 2012). This is a visual lexical decision test commonly used to assess language proficiency, in which participants are presented with 60 English words and nonwords one at a time and asked to indicate whether each is an actual English word or not. Additionally, participants completed the Self-Assessment Manikin (SAM). The SAM scale is a non-verbal pictorial assessment technique that measures emotional arousal, valence and dominance associated with a person's reaction to stimuli (Bradley & Lang, 1994). Participants indicated the level of arousal (ranging from calm to aroused) and valence (ranging from positive to negative) they felt after exposure to each word from the study. For this study, we did not include the measure of dominance.

2.1.3 Procedure

Participants completed the study online individually through the data collection platform Prolific. Participants were first directed to a Qualtrics (Qualtrics, Provo, UT) survey link where they provided consent. Half of the participants in each experiment were further asked to consent to video recording the session. These recordings were later viewed by a research assistant to assess compliance with the instructions. Following consent, participants were given general instructions and then asked to download an E-prime Go^{TM} (Psychology Software Tools, Pittsburgh, PA) file, which directed them through the encoding and free recall portion of the study. Once completed, they were redirected back to the Qualtrics page to complete the LexTALE and SAM assessments.

2.1.3.1 Encoding Phase

Participants were presented with one word list at a time (unrelated, related or emotional) and asked to commit as many words as possible to memory for a later test. Each word was presented centrally on the screen in "Calibri", 42-point font, for 3 seconds. To ensure a deep level of processing, participants were asked to rate, during encoding, how frequently they encounter each word on a scale from 1 – 7 (1: never encounter this word; 7: encounter this word several times a day). A 100ms beep played at the end of the 3 seconds, followed by a fixation cross shown in the center of the screen for 500ms. Once all 16 words in a given list were presented, participants engaged in a filler task for 30 seconds in which they were asked to repeatedly say ‘*bap*’ out loud to eliminate any verbal rehearsal. The order of lists presented was counterbalanced across participants and words were presented in random order within each list type.

2.1.3.2 Retrieval Phase

Following the encoding of each list, and once the 30-second filler task was completed, participants were given 2 minutes to freely recall (by typing into a text box) as many words as possible from the most recently presented list. The screen did not advance until the entire 2 minutes had elapsed. The procedure was repeated for all three lists. Following the recall phase for each of the lists, participants engaged in a visual task for 1 minute to minimize carry-over between lists. In this task, participants were asked to indicate whether two complex images were the same or different.

Following completion of encoding and retrieval of all three list types, participants were redirected back to the Qualtrics survey to complete the LexTALE English proficiency test. They also completed the SAM scale for each of the 48 words presented in the study. They were then thanked and shown a feedback letter explaining the study, as well as a link to receive their remuneration.

2.2 Results

2.2.1 Free Recall

IBM SPSS (v. 28.0.1) statistical software was used to conduct statistical analyses. Recall scores were tabulated by counting the number of correctly recalled (typed) words, separately for each list type. Performance was compared in a 2 (Experiment: 1a and 1b, between subjects) x 3 (List type: Unrelated, Related and Emotional, within-subjects) mixed ANOVA[1]. Memory performance is shown in Table 2.

---

[1] There were no significant differences in recall performance across participants who were vs. were not asked to video record themselves during the experiment (all *p*’s > .05)

Table 2

Mean number of words recalled (standard deviation in parentheses) by list type in Experiments 1a and 1b.

	Recall Performance	
Word List	Experiment 1a	Experiment 1b
Unrelated	4.97 (2.18)	5.96 (2.21)
Related	8.8 (2.97)	8.48 (2.10)
Emotional	7.19 (2.57)	8.32 (2.20)

We found a significant main effect of List Type, $F(2, 274) = 104.85$, $MSE = 373.07$, $p < .001$, $\eta_p^2 = 0.43$ and a marginal main effect of Experiment, $F(1, 137) = 3.72$, $MSE = 37.38$, $p = .056$, $\eta_p^2 = 0.09$. There was also a significant Experiment x List Type interaction, $F(2, 274) = 6.26$, $MSE = 22.28$, $p = .002$, $\eta_p^2 = 0.04$. Follow up comparisons within each Experiment revealed that, in Experiment 1a, recall performance differed significantly for all pairs of lists, whereas in Experiment 1b, recall performance for the related and emotional word lists did not significantly differ (see Table 1). However, importantly, the emotional words were recalled significantly better than unrelated words in both experiments (Exp 1a: $t(138) = 5.49$, $SE = 0.40$, $p < .001$ and Exp 1b: $t(138) = 6.30$, $SE = 0.38$, $p < .001$). Similarly, related words were recalled significantly better than unrelated words in both experiments (Exp 1a: $t(138) = 8.70$, $SE = 0.44$, $p < .001$ and Exp 1b: $t(138) = 6.87$, $SE = 0.37$, $p < .001$).

2.2.2 Intrusions and Errors

Next, we examined intrusion data. Intrusions are words that a participant recalls from one word list during the recall test for a different word list (e.g., while doing the recall test for the related word list, they might produce a word from the emotional list). For Experiment 1a, participants produced an average of 0.13 (*SD*

= 0.54), 0.27 (*SD* = 0.51), and 0.44 (*SD* = 0.50) intrusions from the unrelated, related, and emotional word lists respectively. For Experiment 1b, these means were 0.16 (*SD* = 0.52), 0.17 (*SD* = 0.51), and 0.19 (*SD* = 0.49) for unrelated, related and emotional word lists respectively. To examine the intrusions data, we conducted a 2 (Experiment: 1a, 1b) x 3 (List type: Unrelated, Related and Emotional) mixed ANOVA. We found a significant main effect of List type, $F(2, 274) = 4.65$, $MSE = 1.03$, $p = .01$, $\eta_p^2 = 0.03$, whereby there were significantly more intrusions from the emotional lists (during recall of other lists) than intrusions from the unrelated lists (during recall of other lists). There was also a main effect of Experiment, $F(1, 137) = 86.50$, $MSE = 21.57$, $p < .001$, $\eta_p^2 = 0.39$, such that Experiment 1a produced more intrusions overall than Experiment 1b. Finally, there was a significant Experiment x List type interaction, $F(2, 274) = 3.21$, $MSE = 0.71$, $p = .042$, $\eta_p^2 = 0.02$: the difference in intrusions for emotional vs. unrelated words was only found in Experiment 1a.

Finally, we examined errors. Errors are when a participant incorrectly produces a word that was not presented in any of the encoding lists. Average errors in Experiment 1a were 0.13 (*SD* = 0.34), 0.16 (*SD* = 0.37) and 0.21 (*SD* = 0.41) for unrelated, related and emotional word lists respectively. Errors in Experiment 1b were 0.09 (*SD* = 0.29), 0.13 (*SD* = 0.38) and 0.17 (*SD* = 0.33) for unrelated, related and emotional word lists respectively. There were no significant main effects of List Type, $F(2, 274) < 1$, $p = .845$ or Experiment, $F(2, 274) < 1$, $p = .642$, and no interaction $F(2, 274) = 0.54$, $p = .58$.

2.2.3 Relation between proficiency and recall performance

A Pearson correlation analysis was conducted across the entire sample to examine the relationship between language proficiency, as estimated by LexTALE performance, and recall performance. Raw scores from the LexTALE ranged from 58.75% to 100% correct lexical classification (*M* = 92.28, *SD* = 11.05). Before analysis, outliers (±2 SD) were excluded. This resulted in the exclusion of 4 participants. Following their removal, the adjusted range of scores was 89.25% to 100% (*M* = 95.83, *SD* = 2.14). Individuals with higher English proficiency scores had better recall of the related and emotional lists, $r = 0.24$, $p = .006$ and $r = 0.27$, $p = .002$ respectively, but not the unrelated list, $r = 0.04$, $p = .629$.

2.2.4 Relation between arousal and recall performance

The SAM scale was used to assess the arousal and valence individual participants experienced for each word. Ratings by our participants, collapsed across Experiment 1a and 1b, were in line with ratings from the English Lexicon Project (Balota et al., 2007). SAM scale means for arousal were as follows: Unrelated: *M* = 1.97, *SD* = 0.43; Related: *M* = 2.12, *SD* = 0.49 and Emotional: *M* = 4.01, *SD* = 0.26. SAM scale means for valence were as follows: Unrelated: *M* = 3.35, *SD* = 0.32; Related: *M* = 3.23, *SD* = 0.43 and Emotional: *M* = 1.72, *SD* = 0.21. To determine whether word recall was influenced by arousal, we conducted a word-level logistic regression analysis examining the probability of recall given SAM ratings of arousal. The model explained 9% of the variance in recall, Nagelkerke R^2 = .09. The equation for the regression model is:

$$Y = 0.57X_1 - 2.34$$

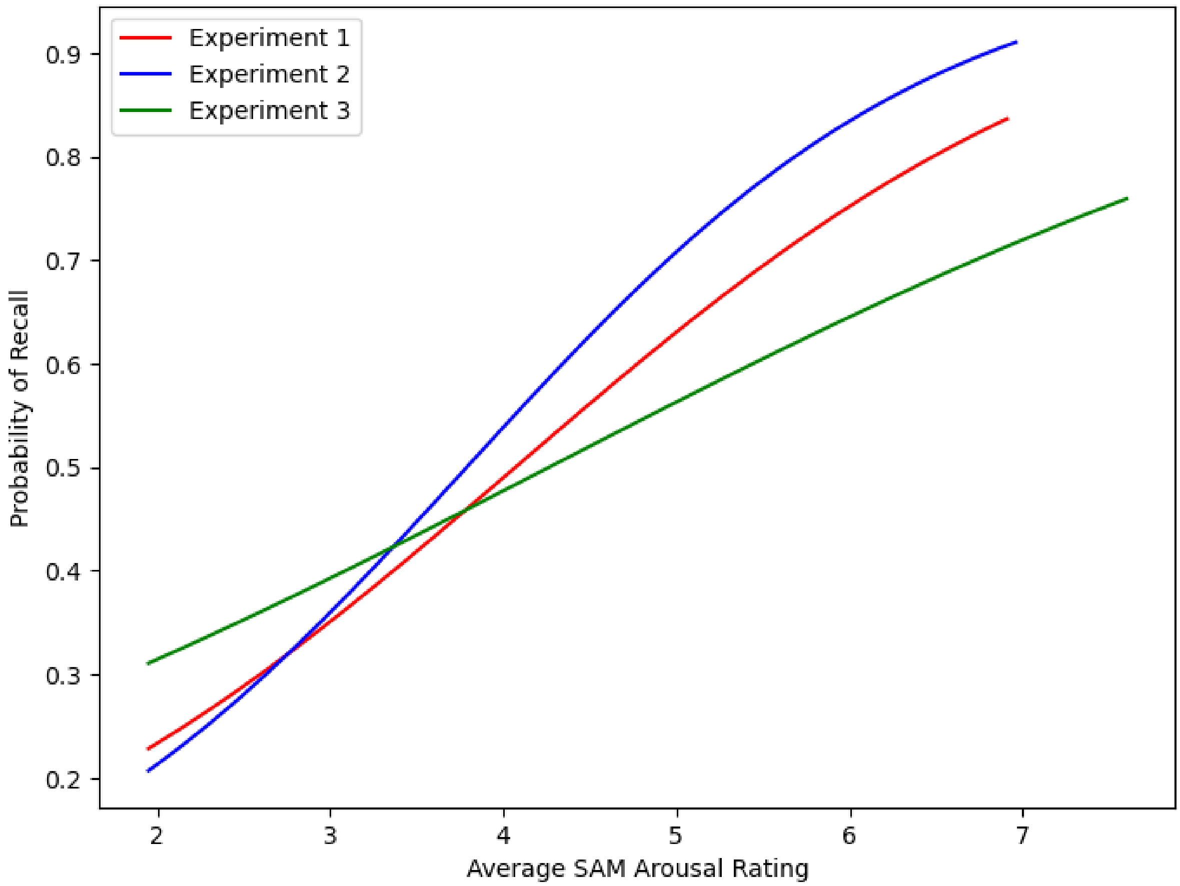
where Y equals the log odds of the model predicting recall and X_1 is the predictor of arousal. This model shows that for every one unit increase in arousal, the odds of recall are increased by 1.77, Wald χ^2 (1, N = 140) = 25.94, $p < .001$, *95% CI* [0.53, 0.62]. To determine the probability of recall given SAM arousal ratings, the equation is transformed to:

$$P(recall) = \frac{1}{1 + e^{-(0.57X_1 - 2.34)}}$$

For example, for the word ‘bomb’ (average arousal rating = 6.21), the model would suggest that the probability of recall would be 77% and for the word ‘pearl’ (average arousal rating = 3.50), the model would suggest that the probability of recall would be 41%. See Figure 1 for the logistic regression curve, showing the relation between SAM arousal ratings and probability of recall, for each of our three experiments. As depicted in Figure 1, there is a positive correlation between SAM arousal ratings and the probability of recall. Thus, while we replicated Talmi and Moscovitch (2004) in finding that related word lists show the same benefit as emotional word lists, this regression shows that arousal predicts memory performance. Based on these results, we suggest that arousal is a key factor underlying enhanced memory for the emotional word lists.

Figure 1

Probability of Recall by Self-Assessment Manikin Arousal Ratings in Experiment 1, 2 and 3



2.3 Discussion

The results of Experiment 1 replicate Talmi and Moscovitch's (2004) work: words from related and emotional lists (both with high relatedness scores) were better recalled than words from unrelated lists (with low relatedness scores). Moreover, words from emotional lists, high in both arousal and relatedness, were no better remembered than words from the related lists, suggesting no additional benefit from arousal. Our regression analysis, however, demonstrates that arousal plays a significant role in predicting memory performance. Moreover, one possible concern about Experiment 1 is the way in which we measured semantic relatedness (using LSA), which may have underestimated the contribution of semantic relatedness. There has been some research regarding the inconsistencies with LSA as a measure of semantic relatedness compared to other approaches (Ensor et al., 2021). For example, Ensor and colleagues (2021) evaluated 2 sets of 13 words for semantic relatedness using a variety of measures, including LSA and WordNet-based techniques (see Wan & Angryk, 2007 for Word-Net based techniques). They found that all measures, except LSA, correctly predicted preferential recall of related words over unrelated words. Therefore, in Experiment 2, we seek to determine whether words from emotionally charged and semantically related lists will show a similar recall benefit when using a different measure of semantic relatedness. We also seek to replicate the results of our regression analysis, which suggest an important role for arousal in predicting recall performance.

Chapter 3
Experiment 2

In Experiment 2, we again examined the relative roles of arousal and semantic relatedness in driving the benefit in memory for emotional words but used a different measure of relatedness – free association. This method is rooted in spreading activation models (Collins & Loftus, 1975; De Deyne et al., 2019; Quillian, 1967). Free Association Norms (FAN; Nelson et al., 2004) provide probability estimates that a given word cues another. In other words, free association index the relative accessibility of related words in a list. For instance, when presented with the word *book*, the most associated response might be *read*.

When we examined our stimuli from Experiment 1a and 1b using the University of South Florida Free Association Norms index (Nelson et al., 2004), we found that although the LSA scores for the emotional word lists were relatively high (Experiment 1a: *LSA* = 0.21, Experiment 1b: *LSA* = 0.24), the free association scores were not (Experiment 1a: *FAN* = 0.09, Experiment 1b: *FAN* = 0.01). We therefore conducted Experiment 2, in which we defined semantic relatedness according to free association strength. Again, Experiment 2 was split into two experiments (2a and 2b). The related and unrelated words in Experiment 2a were the same as in our Experiment 1a (taken from Talmi and Moscovitch (2004)), and those in Experiment 2b were the same as in our Experiment 1b (chosen using the MRC psycholinguistic database). However, we constructed new emotional word lists for Experiments 2a and 2b that were matched to the related word lists using free association norms, rather than LSA.

3.1 Methods

3.1.1 Participants

We collected data from 139 participants (66 males, 73 females) using the online data collection platform Prolific. This sample size was chosen to ensure sufficient power to detect a medium-sized effect in the analysis at an alpha level of .05 based on prior research. Half the participants completed Experiment 2a and the other half completed Experiment 2b. Experiment 2a consisted of 70 participants (31 males, 39 females). Ages ranged from 23 to 67 (M = 41.11, SD = 11.45). Experiment 2b consisted of 69 participants (30 males, 39 females). Ages ranged from 23 to 71 (M = 40.31, SD = 11.13). Participants received remuneration at a rate of £6 per hour. All participants reported their highest level of education completed as an undergraduate degree. All participants reported being monolingual English and raised exclusively in English.

3.1.2 Materials

The unrelated and related word lists were the same as in Experiment 1a and 1b for Experiment 2a and 2b, respectively. Two new lists of emotional words were created for Experiment 2a and 2b using the MRC psycholinguistic database (Wilson, 1988). The difference in this experiment was in the use of the FAN index to create the emotional word lists (and to measure relatedness for the unrelated and related word lists). For each word list, we found the pairwise associations (forward and backward) for every pair of words in the list and averaged these association values to produce an overall average free association score for the entire word list. Thus, words for the new emotional word lists were chosen to result in a high overall free association score according to the FAN index (Nelson et al., 2004). The free association relatedness scores were calculated in the same manner for the other list types.

Within Experiment 2a, one-way ANOVAs confirmed that word lists were matched for the following lexical variables (see Table 2 for means): concreteness: $F(2, 45) = 1.70$, $p = .19$, $\eta_p^2 = 0.07$; familiarity: $F(2, 45) = 1.50$, $p = .23$, $\eta_p^2 = 0.06$; imageability: $F(2, 45) = 1.23$, $p = .30$, $\eta_p^2 = 0.04$; frequency: $F(2, 45) = 0.24$, $p = .79$, $\eta_p^2 = 0.01$ and number of letters: $F(2, 45) = 0.37$, $p = .69$, $\eta_p^2 = 0.02$. Importantly, however, there was a significant difference in valence across word lists $F(2, 45) = 75.24$, $MSE = 0.56$, $p < .001$, $\eta_p^2 = 0.77$, with emotional word lists significantly lower in valence (more negative) than the unrelated ($p < .001$) and related list ($p < .001$). There was also a significant difference in arousal across word lists $F(2, 45) = 36.02$, $MSE = 0.49$, $p < .001$, $\eta_p^2 = 0.62$, with emotional words significantly higher in arousal than the unrelated ($p < .001$) and related ($p < .001$) word lists. Finally, the word lists differed significantly in relatedness, $F(2, 47) = 24.96$, $MSE = 1.43$, $p < .01$, $\eta_p^2 = 0.78$ such that the average FAN score for the unrelated word list ($M_{\text{FAN}} = 0.00$) was significantly lower than the related ($M_{\text{FAN}} = 0.36$), $p < .01$, and emotional ($M_{\text{FAN}} = 0.29$), $p < .01$ lists.

Similarly, within Experiment 2b, word lists were matched for the following lexical variables: familiarity: $F(2, 45) = 2.51$, $p = .09$, $\eta_p^2 = 0.10$; imageability: $F(2, 45) = 2.92$, $p = .06$, $\eta_p^2 = 0.11$; frequency: $F(2, 45) = 0.65$, $p = .52$, $\eta_p^2 = 0.03$ and number of letters: $F(2, 45) = 1.91$, $p = .16$, $\eta_p^2 = 0.08$. Concreteness did differ between word lists, $F(2, 45) = 6.22$, $p < .05$, $\eta_p^2 = 0.22$, such that the average concreteness for emotional words was lower than for the unrelated ($p < .05$) and related ($p < .01$) word lists. This could be due to emotional words being overall lower in concreteness than regular nouns which were used in the unrelated and related word lists. Importantly, a one-way ANOVA revealed a significant difference in valence across word lists $F(2, 45) = 134.36$, $MSE = 0.39$, $p < .001$, $\eta_p^2 = 0.86$, with emotional word lists significantly lower in valence than the unrelated ($p < .001$) and related list ($p < .001$). There was also a significant difference in arousal across word lists $F(2, 45) =$

71.01, *MSE* = 0.31, $p < .001$, $\eta_p^2 = 0.76$, with emotional words significantly higher in arousal than the unrelated ($p < .001$) and related ($p < .001$) word lists. Finally, the word lists differed in relatedness, $F(2, 47) = 36.87$, *MSE* = 0.62, $p < .01$, $\eta_p^2 = 0.77$ such that the average FAN score for the unrelated list ($M_{FAN} = 0.01$) was significantly lower than that for related ($M_{FAN} = 0.31$), $p < .01$ and emotional lists ($M_{FAN} = 0.28$), $p <.01$.

Furthermore, words in Experiment 2a and 2b were matched on several variables. Independent *t* tests indicated no significant differences, collapsed across list types, between experiments in terms of familiarity $t(94) = 0.33$, *SE* = 9.72, $p = .74$, imageability $t(94) = 1.60$, *SE* = 5.81, $p = .11$, number of letters $t(94) = 0.36$, *SE* = 0.21, $p = .72$, valence $t(94) = 0.57$, *SE* = 0.21, $p = .57$, arousal $t(94) = 1.20$, *SE* = 0.16, $p = .23$ and FAN $t(94) = 0.95$, *SE* = 0.62, $p = .35$. However, concreteness, $t(94) = 2.54$, *SE* = 12.42, $p < .05$ and frequency $t(94) = 4.96$, *SE* = 3.45, $p < .001$ did significantly differ between experiments, in that Experiment 2b had lower frequency words, which also naturally tend to be less concrete. See Table 3 for means and standard deviations of lexical variables for Experiment 2.

Table 3

Lexical variables for word lists in Experiment 2

	Experiment 2a			Experiment 2b		
Lexical variables	Unrelated	Related	Emotional	Unrelated	Related	Emotional
Concreteness	512.81 (68.1)	541.63 (65.49)	474 (89.45)	569.19(48.69)	590.13 (31.14)	508.31 (81.19)
Familiarity	520.75 (51.17)	550.19 (48.42)	546.40 (51.87)	556.44 (45.63)	533.94 (44.90)	514.63 (52.82)
Imageability	545 (77.07)	551.75 (55.59)	512.09 (65.80)	553.38 (47.40)	575.13 (35.37)	525.06 (65.41)
Frequency	53.31 (30.93)	56.44 (29.14)	59.60 (16.94)	32.38 (10.90)	36.44 (15.47)	34.56 (26.81)
NLET	5.13 (1.03)	5.06 (1.47)	5.67 (1.66)	5.00 (1.11)	4.81 (0.49)	5.69 (1.47)
Valence	5.51 (0.49)	5.41 (1.00)	2.68 (0.58)	5.54 (0.59)	5.95 (0.49)	2.62 (0.52)
Arousal	3.43 (0.88)	4.14 (0.58)	5.49 (0.55)	3.49 (0.37)	3.32 (0.55)	5.45 (0.65)
FAN	0.00	0.36	0.29	0.01	0.16	0.19

3.1.3 Procedure

The procedure was the same as Experiment 1.

3.2 Results

3.2.1 Free Recall

IBM SPSS (v. 28.0.1) statistical software was used to conduct statistical analyses. Recall scores were tabulated by counting the number of correctly recalled (typed) words, separately for each list type. Performance was compared in a 2 (Experiment: 2a and 2b) x 3 (List type: Unrelated, Related and Emotional) mixed ANOVA[2]. Descriptive statistics are shown in Table 4.

Table 4

Mean number of words recalled (standard deviation in parentheses) by list type in Experiments 2a and 2b

	Recall Performance	
Word List	Experiment 2a	Experiment 2b
Unrelated	4.64 (1.35)	4.48 (1.43)
Related	8.48 (2.20)	8.55 (1.75)
Emotional	8.69 (1.62)	8.26 (2.02)

We found a significant main effect of List Type, $F(2, 274) = 263.32$, $MSE = 752.13$, $p < .001$, $\eta_p^2 = 0.66$, a non-significant effect of Experiment, $F(1, 137) = 2.54$, $MSE = 9.00$, $p = .113$, $\eta_p^2 = 0.02$,

[2] There were no significant differences in recall performance for any of the lists, between participants who were vs. were not asked to video record themselves (all p's > .05)

and no significant Experiment x List Type interaction, $F(2, 274) < 1$, $p = .814$, $\eta_p^2 = 0.02$. Post-hoc tests revealed that, collapsed across Experiments 2a and 2b, recall for the related and emotional word lists was not significantly different ($p = .29$), and recall for the unrelated word list was significantly different from both the related ($p < .001$) and emotional word lists ($p < .001$). See Table 4 for average recall for each list type.

3.2.2 Intrusions and Errors

Moving to the intrusions data, we conducted the same 2 (Experiment: 2a, 2b) x 3 (Intrusion list: Unrelated, Related and Emotional) repeated measures ANOVA. For Experiment 2a, participants produced an average of 0.11 (*SD* = 0.53), 0.19 (*SD* = 0.48), and 0.19 (*SD* = 0.63) intrusions from the unrelated, related, and emotional word lists respectively. For Experiment 2b, these means were 0.13 (*SD* = 0.54), 0.16, (*SD* = 0.39), and 0.14 (*SD* = 0.45) for unrelated, related and emotional word lists respectively. We found no significant main effects of List type, $F(2, 274) < 1$, $p = .528$ or Experiment, $F(1, 137) < 1$, $p = .711$, and no interaction $F(2, 274) = 0.14$, $p = .87$.

Error means for Experiment 2a were 0.11 (*SD* = 0.31), 0.10 (*SD* = 0.37) and 0.09 (*SD* = 0.13) for unrelated, related and emotional word lists respectively. The error means for Experiment 2b were 0.06 (*SD* = 0.23), 0.19 (*SD* = 0.23) and 0.10 (*SD* = 0.14) for unrelated, related and emotional word lists respectively. There were no significant main effects of List type, $F(2, 274) < 1$, $p = .971$ or Experiment $F(1, 137) < 1$, $p = .524$, and no interaction $F(2, 274) = 0.52$, $p = .81$.

3.2.3 Relation between proficiency and recall

A Pearson correlation analysis was conducted to examine the relationship between language proficiency (LexTALE scores) and recall performance. The LexTALE scores ranged from 75.25% to 100% (*M* = 95.37, *SD* = 9.57). Before analysis, any outliers (±2 SD) from the mean score were excluded. This resulted in the exclusion of 2 participants, across the entire sample. Following removal of these outliers, the adjusted range was 87.50% to 100% (*M* = 96.34, *SD* = 2.63). We found a relationship between recall of the emotional list and proficiency scores, *r* = 0.27, *p* = .002. There was no significant relationship between recall of the unrelated and related lists and proficiency, *r* = 0.08, *p* = .347 and *r* = 0.00, *p* = .955 respectively.

3.2.4 Relation between arousal and recall performance

The SAM scale was used to assess participants' ratings of arousal and valence for each word, as in Experiment 1. Ratings by our participants were in line with ratings from the English Lexicon Project (Balota et al., 2007): collapsed across Experiments 2a and 2b, SAM scale means for arousal were as follows: Unrelated: *M* = 1.84, *SD* = 0.41; Related: *M* = 1.96, *SD* = 0.34 and Emotional: *M* = 4.12, *SD* = 0.35. SAM scale means for valence were as follows: Unrelated: *M* = 3.45, *SD* = 0.31; Related: *M* = 3.33, *SD* = 0.57 and Emotional: *M* = 1.65, *SD* = 0.25.

Additionally, we conducted a word-level logistic regression analysis collapsed across Experiments 2a and 2b to determine the probability of recall given SAM ratings of arousal. This model explained 10% of the variance in outcomes, Nagelkerke R^2 = .10. The equation for the regression model is:

$$Y = 0.73X_1 - 2.77$$

where *Y* equals the log odds of the model predicting recall and X_1 is the predictor of arousal. This model shows that for every one unit increase in arousal, the odds of recall is increased by 2.08, Wald χ^2 (1, *N* = 139) = 27.38, *p* < .001, *95% CI* [0.68,0.78]. To determine the probability of recall given SAM arousal ratings, the equation is transformed to:

$$P(recall) = \frac{1}{1 + e^{-(0.73X_1 - 2.77)}}$$

Thus, for the word '*bomb*' (average arousal rating = 5.94), the model would suggest that the probability of recall would be 83% and for the word 'pearl' (average arousal rating = 3.91), the model would suggest that the probability of recall would be 52%. See Figure 1 for the logistic regression curve in each of our 3 experiments.

3.3 Discussion

Once again, the results of Experiment 2 replicate those found in Talmi and Moscovitch (2004), demonstrating that recall of emotional and related word lists did not differ significantly. This finding held even after we changed the method of measuring semantic relatedness. In Experiment 1, we used LSA, which measures semantic relatedness by analyzing patterns of word co-occurrence in large text corpora, identifying latent structures that reflect underlying semantic relationships. In contrast, in Experiment 2, we used Free Association Norms to assess semantic relatedness based on human word associations, capturing how individuals naturally connect words based on experiences. Importantly however, although we find no difference in recall for emotional and related lists in either experiment (consistent with the claims of Talmi and Moscovitch, 2004), the results of our regression analyses in Experiment 2 again highlight the significant role of arousal in memory for emotional word lists.

Chapter 4
Experiment 3

In both Experiment 1 and 2, emotional word lists were matched in relatedness (by LSA in Experiment 1 and by free association in Experiment 2) to the related lists. In both experiments, memory of the emotional and of the related lists was higher than for unrelated lists. However, because emotional word lists were characterized by both high arousal and high relatedness, it was not possible to determine which factor primarily drove the memory advantage. If relatedness is critical for the EEM benefit, then reducing semantic relatedness should also diminish the memory benefit. This is what we tested in Experiment 3. To do so, we manipulated the emotional word lists to maintain high levels of arousal, but lower relatedness. This was accomplished by selecting a mix of high arousal emotional words, such that half were positive and half were negative in valence. This effectively reduced the semantic relatedness of the emotional word lists while keeping arousal high.

4.1 Methods

4.1.1 Participants

We collected data from 139 participants (70 males, 69 females) using Prolific. This sample size was chosen to ensure sufficient power to detect a medium-sized effect in the analysis at an alpha level of .05 based on prior research. Half the participants completed Experiment 3a and the other half completed Experiment 3b. Experiment 3a consisted of 69 participants (29 males, 40 females). Ages ranged from 24 to 61 (M = 40.35, SD = 10.97). Experiment 3b consisted of 70 participants (33 males, 37 females). Ages ranged from 23 to 68 (M = 39.27, SD = 10.27). Participants received remuneration at a rate of £6 per hour. All participants reported their highest level of education completed as an undergraduate degree. All participants reported being monolingual English and raised exclusively in English.

4.1.2 Materials

Semantic relatedness was estimated using the same method as Experiment 2, the FAN association norms. In Experiment 3a, participants were presented with the unrelated and related word lists from Talmi and Moscovitch (2004), as in Experiments 1a and 2a; in Experiment 3b, participants were presented with the same related and unrelated lists as in Experiments 1b and 2b. The critical difference in Experiment 3 was in the construction of the emotional word lists. In contrast to Experiments 1 and 2, which involved only negatively valenced words, the emotional lists in Experiment 3a and 3b were half positively valenced and half negatively valenced. This served the purpose of reducing the semantic relatedness of the emotional words while maintaining high arousal.

Within Experiment 3a, one-way ANOVAs confirmed that word lists were matched for the following lexical variables (see Table 5 for means): familiarity: $F(2, 45) = 2.12$, $p = .13$, $\eta_p^2 = 0.09$; imageability: $F(2, 45) = 1.22$, $p = .31$, $\eta_p^2 = 0.03$; frequency: $F(2, 45) = 1.33$, $p = .28$, $\eta_p^2 = 0.06$ and number of letters: $F(2, 45) = 1.38$, $p = .26$, $\eta_p^2 = 0.06$. Concreteness did differ between word lists, $F(2, 45) = 3.30$, $p < .05$, $\eta_p^2 = 0.13$, such that the average concreteness for emotional words was lower than for the unrelated ($p < .05$) and related ($p < .05$) words. Again, this could be due to emotional words being overall lower in concreteness than regular nouns which were used in the unrelated and related word lists. Importantly, however (and in contrast to Experiments 1 and 2) there was no significant difference in valence across word lists, $F(2, 45) = 0.70$, $p = 0.50$, $\eta_p^2 = 0.03$. At the same time, there was still a significant difference in arousal across word lists, $F(2, 45) = 66.29$, $MSE = 0.49$, $p < .001$, $\eta_p^2 = 0.75$, with emotional words significantly higher in arousal than the unrelated ($p < .001$) and related ($p < .001$) word lists. Finally, the word lists differed significantly in relatedness, $F(2, 47) = 35.75$, $MSE = 0.75$, $p < .01$, $\eta_p^2 = 0.69$ such that the FAN scores for the unrelated word list (M_{FAN}

= 0.01) and emotional word list (M_{FAN} = 0.02) were significantly lower than for the related (M_{FAN} = 0.25), $p < .01$.

Similarly, within Experiment 3b, word lists were matched for the following lexical variables: concreteness: $F(2, 45) = 2.91$, $p = .06$, $\eta_p^2 = 0.11$; familiarity: $F(2, 45) = 0.84$, $p = .44$, $\eta_p^2 = 0.04$; imageability: $F(2, 45) = 1.09$, $p = .35$, $\eta_p^2 = 0.05$; frequency: $F(2, 45) = 0.93$, $p = .40$, $\eta_p^2 = 0.04$ and number of letters: $F(2, 45) = 0.46$, $p = .63$, $\eta_p^2 = 0.02$. Importantly, a one-way ANOVA revealed no significant difference in valence across word lists, $F(2, 45) = 2.72$, $MSE = 2.51$, $p = .08$, $\eta^2 = 0.11$. There was a significant difference in arousal across word lists, $F(2, 45) = 66.29$, $MSE = 0.49$, $p < .001$, $\eta_p^2 = 0.80$, with emotional words significantly higher in arousal than the unrelated ($p < .001$) and related ($p < .001$) word lists. Finally, the word lists differed in relatedness, $F(2, 47) = 32.97$, $MSE = 1.05$, $p < .01$, $\eta_p^2 = 0.59$ such that the FAN scores for the unrelated word list (M_{FAN} = 0.00) and emotional word list (M_{FAN} = 0.01) were significantly lower than for the related (M_{FAN} = 0.16), $p < .01$.

Additionally, words across Experiments 3a and 3b were matched on several variables. Independent *t* tests indicated no significant differences between experiments in terms of familiarity, $t(94) = 0.13$, $SE = 7.90$, $p = .90$, imageability $t(94) = 1.64$, $SE = 6.99$, $p = .10$, number of letters $t(94) = 0.18$, $SE = 0.17$, $p = .86$, valence $t(94) = 0.35$, $SE = 0.23$, $p = .73$, arousal $t(94) = 1.18$, $SE = 0.19$, $p = .24$ and FAN $t(94) = 1.23$, $SE = 1.32$, $p = .22$ However, concreteness, $t(94) = 3.42$, $SE = 10.4$, $p < .001$ and frequency $t(94) = 5.55$, $SE = 3.12$, $p < .001$ did significantly differ between experiments, in that Experiment 3b had lower frequency words, which also tend to be less concrete. See Table 5 for means and standard deviations of lexical variables for Experiment 3.

Table 5

Lexical variables for word lists in Experiment 3

	Experiment 3a			Experiment 3b		
Lexical variables	Unrelated	Related	Emotional	Unrelated	Related	Emotional
Concreteness	512.81 (68.1)	541.63 (65.49)	469.40 (71.72)	569.19(48.69)	590.13 (31.14)	531.56 (85.32)
Familiarity	520.75 (51.17)	550.19 (48.42)	550.47 (40.06)	556.44 (45.63)	533.94 (44.90)	536.06 (50.87)
Imageability	545 (77.07)	551.75 (55.59)	524 (54.21)	553.38 (47.40)	575.13 (35.37)	553.25 (42.71)
Frequency	53.31 (30.93)	56.44 (29.14)	68.20 (31.87)	32.38 (10.90)	36.44 (15.47)	31.13 (16.90)
NLET	5.13 (1.03)	5.06 (1.47)	4.60 (1.16)	5.00 (1.11)	4.81 (0.49)	5.06 (0.85)
Valence	5.51 (0.49)	5.41 (1.00)	5.04 (2.44)	5.54 (0.59)	5.95 (0.49)	5.44 (0.83)
Arousal	3.43 (0.88)	4.14 (0.58)	6.17 (0.62)	3.49 (0.37)	3.32 (0.55)	5.95 (0.98)
FAN	0.01	0.41	0.02	0.00	0.16	0.01

4.1.3 Procedure

The procedure was the same as Experiment 1 and 2.

4.2 Results

4.2.1 Free Recall

IBM SPSS (28.0.1) statistical software was used to conduct statistical analyses. Descriptive statistics are shown in Table 6. Performance was compared in a 2 (Experiment: 3a and 3b) × 3 (List type: Unrelated, Related and Emotional) mixed ANOVA[3].

Table 6

Mean number of words recalled (standard deviation in parentheses) by list type in Experiments 3a and 3b

	Recall Performance	
Word List	Experiment 3a	Experiment 3b
Related	8.77 (3.15)	8.61 (2.03)
Emotional	8.04 (2.26)	8.16 (2.19)
Unrelated	4.84 (1.95)	4.75 (1.99)

We found a significant main effect of List Type, $F(2, 274) = 5.39$, $MSE = 23.79$, $p = .005$, $\eta_p^2 = 0.04$. Post-hoc tests revealed that recall for the unrelated word lists was significantly lower than recall for both related ($p < .01$) and emotional lists ($p < .01$). Recall for related and emotional word

[3] Again, there were no significant differences in recall performance in participants who were vs. were not asked to video record themselves (all p's > .05)

lists were not significantly different from each other (p = .103). There was no significant effect of Experiment, $F(1, 137) < 1$, p = .863 or Experiment x List Type interaction, $F(2, 274) < 1$, p = .456. See Table 6.

4.2.2 Intrusions and Errors

Moving to the intrusions data, we conducted the same 2 (Experiment: 3a, 3b) × 3 (Intrusion list: Unrelated, Related and Emotional) repeated measures ANOVA. For Experiment 3a, participants produced an average of 0.14 (SD = 0.50), 0.30 (SD = 0.36), and 0.30 (SD = 0.42) intrusions from the unrelated, related, and emotional word lists respectively. For Experiment 3b, these means were 0.12 (SD = 0.44), 0.17, (SD = 0.48), and 0.19 (SD = 0.44) for unrelated, related and emotional word lists respectively. We found no significant main effects of List type $F(2, 274) < 1$, p = .39, or Experiment $F(1, 137) < 1$, p = .17, and no interaction $F(2, 274) = 1.36$, p = .24, on number of intrusions.

Error means for Experiment 3a were 0.10 (SD = 0.23), 0.19 (SD = 0.23) and 0.14 (SD = 0.13) for unrelated, related and emotional word lists respectively. The error means for Experiment 3b were 0.04 (SD = 0.12), 0.06 (SD = 0.21) and 0.12 (SD = 0.12) for unrelated, related and emotional word lists respectively. There were no significant effects of List, $F(2, 274) < 1$, p = .966 or Experiment $F(2, 274) < 1$, p = .113, and no interaction $F(2, 274) = 0.34$, p = .85.

4.2.3 Relation between proficiency and recall

A Pearson correlation analysis was conducted to examine the relationship between language proficiency (LexTALE scores) and recall performance. Scores ranged from 72.50% to 100% (M = 91.69, SD = 10.97). Before analysis, any outliers (±2 SD) from the mean proficiency score were excluded. This resulted in the exclusion of 3 participants, across the entire sample. Following the removal of these outliers, the adjusted range was 88.25% to 100% (M = 93.13, SD = 2.98). We found a relationship between proficiency scores and recall for the emotional lists and the related lists, r =

0.31, $p < .001$ and $r = 0.21$, $p = .015$. There was no significant relationship between proficiency scores and recall of words on the unrelated list, $r = 0.05$, $p = .564$.

4.2.4 Relation between arousal and recall performance

The SAM scale was used to assess participants' ratings of arousal and valence for each word. Ratings by our participants were in line with ratings from the English Lexicon Project (Balota et al., 2007): collapsed across Experiments 3a and 3b, SAM scale means for arousal were as follows: Unrelated: $M = 1.88$, $SD = 0.37$; Related: $M = 1.93$, $SD = 0.24$ and Emotional: $M = 4.23$, $SD = 0.30$. SAM scale means for valence were as follows: Unrelated: $M = 3.55$, $SD = 0.32$; Related: $M = 3.68$, $SD = 0.46$ and Emotional: $M = 1.65$, $SD = 0.34$.

We conducted a word-level logistic regression analysis to determine the probability of recall given SAM ratings of arousal. This model explained approximately 5% of the variance in outcomes, Nagelkerke $R^2 = .05$. The equation for the regression model is:

$$Y = 0.34X_1 - 1.47$$

where Y equals the log odds of the model predicting recall and X_1 is the predictor of arousal. This model shows that for every one unit increase in arousal, the odds of recall is increased by 1.40, Wald χ^2 (1, $N = 139$) = 17.98, $p < .01$, *95% CI* [0.31, 0.38]. To determine the probability of recall given SAM arousal ratings, the equation is transformed to:

$$P(recall) = \frac{1}{1 + e^{-(0.34X_1 - 1.47)}}$$

Thus, for the word bomb (average arousal rating = 5.63) the model would suggest that the probability of recall would be 69% and for the word 'pearl' (average arousal rating = 3.11) the model would suggest that the probability of recall would be 40%.

4.3 Discussion

The results from Experiment 3 provide strong support for an arousal-based mechanism underlying the EEM. Even though the emotional word list had lower average semantic relatedness—due to the inclusion of both positive and negative valenced words, the emotional word list still had a memory benefit of the same magnitude as the semantically related list, as in Experiments 1 and 2. Such a pattern cannot be accounted for by a relatedness account. Our findings therefore suggest, in line with Cahill and McGaugh's (1998) account, that arousal is the primary factor driving memory enhancement for emotional words. This claim is bolstered by the consistent positive relationship between arousal and predicted recall in all three of our experiments (see Figure 1).

Chapter 5
General Discussion

We compared an arousal and semantic relatedness account of immediate emotional enhancement in memory for word lists. In three experiments, participants were asked to encode lists composed of words that were either unrelated, related, or emotional. In Experiments 1 and 2, latent semantic analysis (Experiment 1) and free association norms (Experiment 2) were used to match the relatedness of words in the emotional (negative) and related lists. Memory for related and emotional words was significantly higher than for words in the unrelated lists. In Experiment 3, we manipulated the emotional lists to be mixed-valence, thereby lowering semantic relatedness but keeping arousal high. Despite lowering relatedness of the emotional word lists, we found that memory for related and emotional word lists was again of similar magnitude, with memory for both significantly higher than for unrelated lists. Importantly, participant ratings of arousal were found to predict recall. Overall, our findings suggest that arousal, more than semantic relatedness, explains why memory is better for lists of emotional words.

Experiments 1 and 2 successfully replicated the findings of Talmi and Moscovitch (2004), showing that emotional and related word lists were comparable in memory performance. Additionally, we extended their findings to other word lists and measures of semantic relatedness, demonstrating good replicability and the generalizability of their results across different methods. Across all three experiments, we consistently observed better memory for emotional and related word lists compared to the unrelated word lists. However, the results of Experiment 3 conflict with a pure semantic relatedness account of the immediate EEM. In Experiment 3, we manipulated the emotional word list to include an equal proportion of positive and negative valence words. Despite reducing semantic relatedness, the EEM benefit persisted. Importantly, this finding highlights that the EEM cannot solely be driven by semantic relatedness, contrary to Talmi and Moscovitch's (2004) proposal. Moreover, in all 3 experiments, we found that arousal predicted recall. Together, these findings suggest that an arousal mechanism remains the best account for the immediate EEM effect.

5.1 Arousal vs. Semantic Relatedness

Our study provides two strong pieces of support for arousal and attention as the driving factors underlying the immediate EEM effect. First, the inclusion of a mixed-valence emotional word list in Experiment 3 reduced the interrelatedness of the emotional words, while maintaining high levels of arousal. To our knowledge, this is the first study to explore the EEM mechanism using a combination of both positive and negative valence words. Despite this manipulation, the memory advantage for emotional words persisted, which aligns with the view that the immediate enhanced recall of emotional words may stem from heightened arousal and attention. Second, the SAM ratings across all three experiments indicated that participants' levels of arousal for each word significantly predicted recall performance, with the probability of recall increasing with increasing arousal ratings.

Together, our findings are consistent with suggestions that the immediate memory benefit provided by emotional stimuli can be attributed to an increase in arousal and amygdala activity, and not semantic relatedness. The present findings align with previous evidence suggesting that emotional stimuli receive heightened attentional processing during encoding compared to neutral stimuli (Fox et al., 2001; MacKay et al., 2004; Schimmack & Derryberry, 2005; Williams et al., 1996). This enhanced attention is thought to facilitate deeper perceptual engagement, as indicated by increased activation in regions such as the anterior temporal and extrastriate visual cortex when viewing emotional material (Bradley et al., 2003; Mourão-Miranda et al., 2003). Given the well-established role of attention in supporting memory formation (Craik et al., 1996; Smith, 1895), it is plausible that emotional stimuli are better remembered in part because they draw more attention at encoding

(Hamann, 2001; Kensinger & Corkin, 2003). Furthermore, the observed effects align with prior fMRI research showing that emotional content enhances both attention and memory through increases in amygdala activity, consistent with findings that the immediate EEM benefit is also associated with increased amygdala activity (MacKay et al., 2004; Mickley Steinmetz et al., 2010; Pratto & John, 1991). Taken together, these findings support the interpretation that elevated attentional engagement with emotional stimuli may reflect underlying increases in arousal, which in turn contribute to enhanced memory performance.

That said, we consistently observed clear benefits of semantic relatedness for the related word lists in our experiments. It is also worth noting that we observed significant correlations between LexTALE performance, a proxy for English proficiency, and recall of the emotional and semantic word lists. In contrast, there were no correlations between LexTALE score and recall for unrelated word lists in any experiment. Previous work suggests that as proficiency in a language increases, so too does the strength of word connections in the spreading activation network (Collins & Loftus, 1975; Quillian, 1967). Therefore, the increase in recall performance along with increases in proficiency suggest that relatedness may have influenced memory for the emotional lists (in addition to the related lists). However, given that increased proficiency may also enhance the efficiency of processing emotional words (Sharif & Mahmood, 2023), as well as the fact that our sample consisted of English monolinguals with a limited range of proficiencies, this possibility will need to be explored in future work.

5.2 Additive effects of semantic relatedness and arousal?

Although words in our emotional lists in Experiments 1 and 2 were high in both semantic relatedness and arousal, their benefit to memory was not additive – rather, they showed the same benefit relative to neutral unrelated words as the related word lists did. This is not a result of an inherent memory ceiling effect. Previous work using lists of the same length as in our study (16 words) indicates that average recall is around 11 out of 16 (68% recall) (Graves et al., 2021). In our work, the overall average recall for emotional words was 8.11 out of 16 (51% recall) and related words was 8.62 out of 16 (54% recall).

Previous work has shown that combining multiple factors known to benefit memory can produce an additive effect such that including both factors enhances memory over and above using a single beneficial factor (Fawcett et al., 2012; MacLeod et al., 2010). Our lack of an additive effect is unlikely to be a result of an inadequate relatedness measure given the similarity in recall of emotional lists in Experiments 1 and 2, which used two different measures of relatedness. Instead, the lack of an additive effect may suggest that arousal and relatedness do not independently enhance memory but rather interact.

On the other hand, an additive effect was observed in a study by Buchanan and colleagues (2006). In addition to related, emotional and unrelated word lists, that study included a fourth set of word lists with taboo words (sexually explicit words and profanities). As in our study, memory for the list of emotional words was no better than recall of their related word list. Interestingly, however, they did find a significant difference between memory for taboo words and memory for all other word lists. These taboo words were both semantically related and extremely high in arousal. Based on these results, the authors suggest the effect of arousal and relatedness are additive. However, other work (Janschewitz, 2008) suggests that taboo words are used much less frequently than are emotional words, and this impacts memory. This relates to distinctiveness theory, as taboo stimuli are typically much less frequent and have unique attributes that stand out relative to other items. There is strong evidence that distinctiveness improves memory (Hunt & McDaniel, 1993, Schmidt, 1991). Additionally, taboo and sexually explicit words evoke strong social and cultural associations, triggering intense personal reactions after exposure. Whether arousal and semantic relatedness are additive requires further investigation.

5.3 Future Directions

Although we did not directly manipulate attention in this series of studies, we made a reasonable statement that arousal and attention are linked. A future study could look to directly manipulate attentional states during encoding, similar to what was done in Yeung & Fernandes (2021). In that study, the authors investigated how divided attention during encoding or retrieval affects memory for emotional words. They found that divided attention impaired memory for negative emotional words, eliminating their typical memory advantage, while positive words were unaffected. If it is true that arousal and attention are the true driving force of the immediate EEM, it would make sense that participants in the divided attention condition show a weakened EEM effect. This can further be extended to an eye pupil study to directly measure attentional states during encoding. Additionally, the measure of arousal used in this study was a self-rated assessment that has been shown to be well correlated with physiological responses. However, in order to deepen the understanding of how arousal and amygdala activity relate to emotional memory, a more physiological approach using skin conductance or heart rate could be applied.

Moreover, a plausible explanation for the comparable recall of words from the emotional list and neutral words could stem from the timing of the recall test – an immediate test conducted shortly after encoding. As discussed, there may be greater impact of arousal mechanisms on memory after the process of consolidation has occurred. To explore this further, future experiments should consider delaying the recall test by hours or even days as we know from previous research that sleep can be an essential part of the memory consolidation process.

Finally, our results may also be limited to word stimuli. Research suggests that visual stimuli such as pictures or videos elicit more intense arousal responses, both positive and negative, compared to words alone. Differential processing of emotional stimuli under these conditions could accentuate any observed differences. Therefore, future studies should investigate whether the EEM effect is influenced more by arousal or semantic relatedness when using emotional pictures or videos.

5.4 Conclusion

Across three experiments, we provide evidence supporting the idea that the immediate EEM is better explained by the attentional arousal-based account than by a semantic relatedness account (Talmi & Moscovitch, 2004). By employing two distinct methods for estimating semantic relatedness, we further demonstrated that, for emotional words, arousal and semantic relatedness do not work as entirely independent factors. Specifically, we did not observe an additive benefit when arousal and relatedness were combined. Such a finding suggests that, at least at the levels of arousal we examined, relatedness and arousal are interactive. The nature of this interaction remains unclear, leaving room for future research to further investigate and disentangle their relationship.

Additionally, the analysis of the SAM arousal ratings provides strong evidence for the role of arousal in enhancing memory recall. Our analyses revealed a significant positive relationship between arousal levels and the probability of recall, indicating that as arousal increased, the likelihood of recalling words also increased. Importantly, this finding aligns with theories proposing that emotionally arousing stimuli receive prioritized processing, which strengthens their encoding and subsequent retrieval. These results deepen our understanding of the underlying mechanisms that contribute to the EEM effect, highlighting the pivotal role of arousal in enhancing memory performance.

References

Aubé, W., Peretz, I., & Armony, J. L. (2013). The effects of emotion on memory for music and vocalisations. *Memory (Hove, England)*, *21*(8), 981–990. https://doi.org/10.1080/09658211.2013.770871

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 576–587. https://doi.org/10.1037/0278-7393.6.5.576

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barnhardt, T. M., Choi, H., Gerkens, D. R., & Smith, S. M. (2006). Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. *Journal of Memory and Language*, *55*(2), 213–231. https://doi.org/10.1016/j.jml.2006.04.003

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science (New York, N.Y.)*, *269*(5227), 1115–1118. https://doi.org/10.1126/science.7652558

Blake, T. M., Varnhagen, C. K., & Parent, M. B. (2001). Emotionally arousing pictures increase blood glucose levels and enhance recall. *Neurobiology of Learning and Memory*, *75*(3), 262–273. https://doi.org/10.1006/nlme.2000.3973

Bookbinder, S. H., & Brainerd, C. J. (2017). Emotionally negative pictures enhance gist memory. *Emotion*, *17*(1), 102–119. https://doi.org/10.1037/emo0000171

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. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-Trace Theory and children's false memories. *Journal of Experimental Child Psychology*, *71*(2), 81–129. https://doi.org/10.1006/jecp.1998.2464

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*(1), 73–99. https://doi.org/10.1016/0010-0277(77)90018-X

Buchanan, T. W., Etzel, J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic arousal and semantic relatedness on memory for emotional words. *International Journal of Psychophysiology*, *61*(1), 26–33. https://doi.org/10.1016/j.ijpsycho.2005.10.022

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition*, *20*(3), 277–290. https://doi.org/10.3758/BF03199665

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, *4*(4), 410–421. https://doi.org/10.1006/ccog.1995.1048

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*(7), 294–299. https://doi.org/10.1016/s0166-2236(97)01214-9

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

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

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

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

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review*, *14*(3), 423–429. https://doi.org/10.3758/BF03194083

Ensor, T. M., MacMillan, M. B., Neath, I., & Surprenant, A. M. (2021). Calculating semantic relatedness of lists of nouns using WordNet path length. *Behavior Research Methods*, *53*(6), 2430–2438. https://doi.org/10.3758/s13428-021-01570-0

Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin*, *83*(3), 389–404. https://doi.org/10.1037/0033-2909.83.3.389

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory (Hove, England)*, *20*(7), 655–666. https://doi.org/10.1080/09658211.2012.693510

Finkenauer, C., Luminet, O., Gisle, L., el-Ahmadi, A., van der Linden, M., & Philippot, P. (1998). Flashbulb memories and the underlying mechanisms of their formation: Toward an emotional-integrative model. *Memory & Cognition*, *26*(3), 516–531. https://doi.org/10.3758/bf03201160

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology. General*, *130*(4), 681–700.

Graves, L. V., Drozdick ,Lisa, Courville ,Troy, Farrer ,Thomas J., Gilbert ,Paul E., & and Delis, D. C. (2021). Cohort differences on the CVLT-II and CVLT3: Evidence of a negative Flynn effect on the attention/working memory and learning trials. *The Clinical Neuropsychologist*, *35*(3), 615–632. https://doi.org/10.1080/13854046.2019.1699605

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394–400. https://doi.org/10.1016/s1364-6613(00)01707-1

Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, *2*(3), 289–293. https://doi.org/10.1038/6404

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*(4), 421–445. https://doi.org/10.1006/jmla.1993.1023

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods*, *40*(4), 1065–1074. https://doi.org/10.3758/BRM.40.4.1065

Kapp, B. S., Frysinger, R. C., Gallagher, M., & Haselton, J. R. (1979). Amygdala central nucleus lesions: Effect on heart rate conditioning in the rabbit. *Physiology & Behavior*, *23*(6), 1109–1117. https://doi.org/10.1016/0031-9384(79)90304-4

Kazanas, S. A., & Altarriba, J. (2015). The automatic activation of emotion and emotion-laden words: Evidence from a masked and unmasked priming paradigm. *The American Journal of Psychology*, *128*(3), 323–336. https://doi.org/10.5406/amerjpsyc.128.3.0323

Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences*, *15*(4), 241–251. https://doi.org/10.1515/revneuro.2004.15.4.241

Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science*, *16*(4), 213–218. https://doi.org/10.1111/j.1467-8721.2007.00506.x

Kensinger, E. A., Brierley, B., Medford, N., Growdon, J. H., & Corkin, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion (Washington, D.C.)*, *2*(2), 118–134. https://doi.org/10.1037/1528-3542.2.2.118

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

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065–1081. https://doi.org/10.1037/a0035669

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, *9*(6), 490–493. https://doi.org/10.1111/1467-9280.00090

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

LeDoux, J. E. (1994). Emotion, Memory and the Brain. *Scientific American*, *270*(6), 50–57.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, *32*(3), 474–488. https://doi.org/10.3758/bf03195840

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Majerus, S., & D'Argembeau, A. (2011). Verbal short-term memory reflects the organization of long-term memory: Further evidence from short-term memory for emotional words. *Journal of Memory and Language*, *64*(2), 181–197. https://doi.org/10.1016/j.jml.2010.10.003

Mandler, G. (1967). Organization and memory. In *The psychology of learning and motivation: I.* (pp. x, 381–x, 381). Academic Press. https://doi.org/10.1016/S0079-7421(08)60516-2

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *2*(1), 33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

Mickley Steinmetz, K. R., Addis, D. R., & Kensinger, E. A. (2010). The effect of arousal on the emotional memory network depends on valence. *NeuroImage*, *53*(1), 318–324. https://doi.org/10.1016/j.neuroimage.2010.06.015

Mourão-Miranda, J., Volchan, E., Moll, J., de Oliveira-Souza, R., Oliveira, L., Bramati, I., Gattass, R., & Pessoa, L. (2003). Contributions of stimulus valence and arousal to visual activation during emotional perception. *NeuroImage*, *20*(4), 1955–1963. https://doi.org/10.1016/j.neuroimage.2003.08.011

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Ozawa, S., Nakatani, H., Miyauchi, C. M., Hiraki, K., & Okanoya, K. (2022). Synergistic effects of disgust and anger on amygdala activation while recalling memories of interpersonal stress: An fMRI study. *International Journal of Psychophysiology*, *182*, 39–46. https://doi.org/10.1016/j.ijpsycho.2022.09.008

Pillemer, D. B. (2003). Directive functions of autobiographical memory: The guiding power of the specific episode. *Memory (Hove, England)*, *11*(2), 193–202. https://doi.org/10.1080/741938208

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology Section A*, *48*(2), 384–404. https://doi.org/10.1080/14640749508401396

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. Journal of Personality and Social Psychology, 61(3), 380–391. https://doi.org/10.1037/0022-3514.61.3.380

Rapaport, D. (1943). Emotions and memory. *Psychological Review*, *50*(2), 234–243. https://doi.org/10.1037/h0055721

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Ross Quillian, M. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430.

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, *14*(1), 79–94. https://doi.org/10.3758/BF03209231

Schimmack, U., & Derryberry, D. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion (Washington, D.C.)*, *5*(1), 55–66. https://doi.org/10.1037/1528-3542.5.1.55

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, *19*(6), 523–542. https://doi.org/10.3758/BF03197149

Sharif, H., & Mahmood, S. (2023). Emotional processing in bilinguals: A systematic review aimed at identifying future trends in neurolinguistics. *Humanities and Social Sciences Communications*, *10*(1), 1–12. https://doi.org/10.1057/s41599-023-01926-1

Smith, W. G. (1895). The relation of attention to memory. *Mind*, *IV*(13), 47–73. https://doi.org/10.1093/mind/IV.13.47

Stricker, J. L., Brown, G. G., Wixted, J., Baldo, J. V., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test-Second Edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society: JINS*, *8*(3), 425–435. https://doi.org/10.1017/s1355617702813224

Subramanian, R., Shankar, D., Sebe, N., & Melcher, D. (2014). Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes. *Journal of Vision*, *14*(3), 31. https://doi.org/10.1167/14.3.31

Talarico, J. M., Berntsen, D., & Rubin, D. C. (2009). Positive emotions enhance recall of perhipheral details. *Cognition & Emotion*, *23*(2), 380–398. https://doi.org/10.1080/02699930801993999

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*(5), 742–751. https://doi.org/10.3758/BF03195864

Talmi, D., Schimmack, U., Paterson, T., & Moscovitch, M. (2007). The role of attention and relatedness in emotionally enhanced memory. *Emotion (Washington, D.C.)*, *7*(1), 89–102. https://doi.org/10.1037/1528-3542.7.1.89

Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review*, *69*(4), 344–354. https://doi.org/10.1037/h0043150

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, *5*(4), 381–391. https://doi.org/10.1016/S0022-5371(66)80048-8

Wan, S., & Angryk, R. A. (2007). Measuring semantic similarity using wordnet-based context vectors. *2007 IEEE International Conference on Systems, Man and Cybernetics*, 908–913. https://doi.org/10.1109/ICSMC.2007.4413585

Waring, J. D., & Kensinger, E. A. (2011). How emotion leads to selective memory: Neuroimaging evidence. *Neuropsychologia*, *49*(7), 1831–1842. https://doi.org/10.1016/j.neuropsychologia.2011.03.007

Williams, J. M., Mathews, A., & MacLeod, C. (1996). The emotional stroop task and psychopathology. *Psychological Bulletin*, *120*(1), 3–24. https://doi.org/10.1037/0033-2909.120.1.3

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6–10. https://doi.org/10.3758/BF03202594

Yeung, R. C., & Fernandes, M. A. (2021). Divided attention at encoding or retrieval interferes with emotionally enhanced memory for words. *Memory (Hove, England)*, *29*(3), 284–297. https://doi.org/10.1080/09658211.2021.1887896

Zimmerman, C. A., & Kelley, C. M. (2010). "I'll remember this!" Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language*, *62*(3), 240–253. https://doi.org/10.1016/j.jml.2009.11.004

Appendix A
Word Lists

Word List Type	Word
	Exp 1a
Unrelated	basket, letter, fork, water, spray, tool, manner, quiet, paper, elbow, hat, rattle, chair, sphere, fabric, journal
Related	car, emission, fuel, highway, lane, van, travel, wheel, driver, road, engine, brake, vehicle, truck, route, gas
Emotional	hell, scream, failure, witch, liar, kill, war, gang, terror, murder, bomb, cancer, death, fear, pain, hate
	Exp 1b
Unrelated	pearl, fork, lunch, jacket, chin, quarter, trail, citizen, bag, metal, plate, brass, chart, truck, couch, belt
Related	sheep, hay, garden, cheese, chicken, harvest, fruit, rice, milk, egg, soil, ranch, seed, rural, cow, animal
Emotional	poison, jail, bomb, terror, cancer, flood, murder, slave, grave, spider, trash, scar, crime, injury, bullet, blood
	Exp 2a
Unrelated	basket, letter, fork, water, spray, tool, manner, quiet, paper, elbow, hat, rattle, chair, sphere, fabric, journal
Related	car, emission, fuel, highway, lane, van, travel, wheel, driver, road, engine, brake, vehicle, truck, route, gas
Emotional	gun, combat, rifle, nuclear, blood, war, army, fight, enemy, death, military, prisoner, battle, conflict, murder, die
	Exp 2b
Unrelated	pearl, fork, lunch, jacket, chin, quarter, trail, citizen, bag, metal, plate, brass, chart, truck, couch, belt
Related	sheep, hay, garden, cheese, chicken, harvest, fruit, rice, milk, egg, soil, ranch, seed, rural, cow, animal
Emotional	tank, assault, bullet, shotgun, bomb, weapon, genocide, pistol, war, trench, grenade, explode, kill, cannon, missile, riot
	Exp 3a
Unrelated	basket, letter, fork, water, spray, tool, manner, quiet, paper, elbow, hat, rattle, chair, sphere, fabric, journal
Related	car, emission, fuel, highway, lane, van, travel, wheel, driver, road, engine, brake, vehicle, truck, route, gas
Emotional	hell, scream, failure, witch, liar, kill, war, gang, cash, gold, sex, party, puppy, music, naked, money
	Exp 3b
Unrelated	pearl, fork, lunch, jacket, chin, quarter, trail, citizen, bag, metal, plate, brass, chart, truck, couch, belt

Related	sheep, hay, garden, cheese, chicken, harvest, fruit, rice, milk, egg, soil, ranch, seed, rural, cow, animal
Emotional	poison, jail, bomb, terror, cancer, flood, murder, slave, orgasm, lover, hero, lottery, candy, laugh, award, cash