VALENCE, AROUSAL, AND DOMINANCE OF EMOTIONAL CUE WORDS USED IN OVERGENERAL MEMORY RESEARCH

Burcu KAYA KIZILÖZ*

ABSTRACT

People in depression tend to retrieve overly general memories instead of specific events (Overgeneral Memory, Williams et al., 2007). The most commonly used measures in overgeneral memory (OGM) research, the Autobiographical Memory Test (AMT), consists of five positive and five negative cue words. OGM studies often fail to find differences in specificity between positive and negative cues and this could partly result from a methodological issue. An examination of 61 studies using AMT revealed that more than 85 cues were chosen without using a standardized selection method. We explored the characteristics of these AMT cues used in OGM research. Forty university students (20 female) rated the cues for valence, arousal and dominance and filled out depression measures. We found that valence, arousal and dominance ratings were modulated by participants' depression level. Our results suggest a strong need to have a standardized selection method for words used in OGM research.

Key Words: *overgeneral memory, autobiographical memory test, valence, arousal*

Aşırı Genelleme Olgusu Araştırmalarında Kullanılan Duygusal Kelimelerin Değerlik, Uyarılma ve Dominantlık Seviyeleri

ÖZET

Depresyondaki bireyler, belirli olaylardan ziyade çoğunlukla aşırı genel anılar hatırlama eğilimindedir (Aşırı Genelleme Olgusu, Williams ve ark., 2007). Aşırı genelleme olgusu (AGO) araştırmalarında en yaygın kullanılan ölçümlerden biri olan Otobiyografik Bellek Testi (ABT), beş olumlu ve beş olumsuz ipucu kelimeden oluşmaktadır. AGO çalışmalarında olumlu ve olumsuz ipuçları arasında anı özgüllüğü bakımından fark bulunamaması sık karşılaşılan bir durumdur ve bu durum kısmen yöntemsel sorunlardan kaynaklanıyor olabilir. ABT kullanan 61 çalışmanın incelenmesi, 85'ten fazla ipucu kelimenin standart bir seçim yöntemi olmaksızın belirlendiğini ortaya koymuştur. Bu çalışmada, AGO araştırmalarında kullanılan ABT ipuçlarının özellikleri incelenmiştir. Kırk üniversite öğrencisi (20 kadın), bu ipuçlarını değerlik, uyarılma ve baskınlık boyutlarında değerlendirmiş ve depresyon ölçeklerini doldurmuştur. Bulgular, değerlik, uyarılma ve baskınlık puanlarının katılımcıların depresyon düzeylerine göre değiştiğini göstermektedir. Bu sonuçlar, OGM araştırmalarında kullanılan kelimeler için standart bir seçim yöntemine duyulan güçlü gereksinime işaret etmektedir.

Anahtar Kelimeler: *aşırı genelleme olgusu, otobiyografik bellek testi, değerlik, uyarılma*

* Assist. Prof. Dr., Eastern Mediterranean University Faculty of Arts and Sciences Department of Psychology, Gazimağusa / TURKISH REPUBLIC OF NORTHERN CYPRUS, burcu.kaya@emu.edu.tr

Research Article / Araştırma Makalesi

Cite as / Atıf: Kaya Kızılöz, B. (2026). Valence, arousal, and dominance of emotional cue words used in overgeneral memory research. *Uludağ University Faculty of Arts and Sciences Journal of Social Sciences, 27*(51), 867-883. https://dx.doi.org/10.21550/sosbilder.1823560

Sending Date / Gönderim Tarihi: 14 November / Kasım 2026
Acceptance Date / Kabul Tarihi: 16 April / Nisan 2026

Introduction

Overgeneral memory (OGM) is a failure to retrieve memories of specific life events when prompted, and it is one of the most studied topics in the autobiographical memory (ABM) field. Memory specificity refers to the ability to retrieve a single, discrete event from one's past that occurred at a specific time and place and lasted no longer than 24 hours (e.g., The dinner I had with my sister last Tuesday night) (Williams & Broadbent, 1986). In contrast, OGM occurs when an individual retrieves a summary of frequent events (e.g., dinners with my sister) or extended periods (e.g., 'The summer I visited London') instead of a specific instance.

Although the first study on OGM (Williams & Broadbent, 1986) focused on participants with depression, OGM is widely examined in both clinical and non-clinical samples (see Summer, 2012 for a review). Autobiographical Memory Test (AMT), introduced by Williams and Broadbent (1986), is one of the most commonly used measures of memory specificity in both clinical and non-clinical samples. The AMT requires the person to retrieve a specific memory (an event that did not last longer than 24 hours) in response to 5 positive (i.e., *happy, safe, interested, successful, surprised*), and 5 negative (i.e., *sorry, angry, clumsy, hurt, lonely*) emotional cue words. Subsequently, independent coders evaluate the specificity of memories retrieved. The AMT is a widely used method to study OGM: A literature review by Sumner (2012) reported that most studies examining OGM used AMT and/or its variation. Nevertheless, the test has received some critique (Sumner et al., 2013) due to its lack of sensitivity to the OGM phenomenon, especially in the non-clinical samples (Bartoli & Smorti, 2018) or its relatively higher accuracy for less specific information (Griffith et al., 2009; Ros et al., 2018). Despite its prevalence, the AMT's reliance on binary valence categories often results in inconsistent findings across studies, suggesting that valence alone is insufficient to capture the cognitive mechanisms driving overgeneralization. This highlights a critical research gap: the need to identify which specific affective properties of cue words determine whether a search process successfully reaches a specific memory or terminates at a general level. The primary aim of the present study is to provide a preliminary multidimensional evaluation of emotional cue words to demonstrate how additional affective dimensions can enhance the methodological precision of the AMT.

Over the years, several modifications were made to the AMT. For instance, Ricarte-Trives et al. (2014) investigated the impact of different instructions on the responses to AMT. They showed that less specific instructions lead to less specific memories because different instructions can have a different cognitive load. Although some cue characteristics like the levels of cue imageability, familiarity, or frequency (e.g., Ricarte-Trives et al., 2014; Ros et al., 2018; Williams et al., 1999) were examined to compare and control for the cue characteristics better, the emotional properties of these emotional cue words and how they can be perceived differently by individuals remain unexplored. In Table 1[1], we presented 61 studies using AMT as their data collection tool, and we specified whether these studies accounted for the valence ratings in a given sample. A limited number did so and compared it to the original cues provided by Williams and Broadbent (1986). Furthermore, the use of cue types highly varied across studies. We reviewed studies published between 1986 and 2022. The keywords we used for the search were "Overgeneral Memory", "Memory Specificity", "Reduced Memory Specificity", "Autobiographical Memory in/and Depression" and "Autobiographical Memory Test". In these 81 studies we reviewed, the percentage of original AMT cue words (Williams & Broadbent, 1986) were as follows: happy (66%), safe (31%), interested (33%), successful (30%), surprised (30%), sorry (15%), angry (45%), clumsy (20%), hurt (38%) and lonely (57%). Furthermore,

[1] The table is provided in the OSF page due to its extensive size.

although 77 additional cue words were used that were not in the original AMT task, only 34% of the studies explained how they picked the cue words they used. Summaries of cues and cue selection methods of studies using AMT to examine OGM can be found on OSF Repository.

Research demonstrates that higher depression or other pathologies' symptoms are associated with lower specificity (see Hallford et al., 2021, for a review). However, in light of previous research (e.g., Simpson & Sheldon, 2020), we argue for a more precise explanation for the effect of cue type on AMT. Typically, the AMT utilizes positive and negative emotional words to trigger memories. Even though the number of cues used in different studies varies, the average remains around 5 cues per valence category, as in Williams and Broadbent (1986). In most studies, the valence categories of positive and negative cues are combined, as the specificity scores are not usually significantly different between valence categories. Supporting this practice, using confirmatory factor analysis, Griffith et al. (2009) showed that the cues could form one factor. However, researchers, not participants, classify the cues as positive or negative; thus, retrieved memories in these valence categories may not reflect the actual valence of the event as perceived by the participant. The failure to assess and control for cue interpretation is a question of concern (Ricarte-Trives et al., 2014). It is important to note that most of the OGM studies sample depressed individuals, which may result in a bias toward interpreting/evaluating the negative stimuli as more negative and positive stimuli as less positive (Gollan et al., 2016).

1. Autobiographical Memory and Cue Characteristics

The retrieval of autobiographical memories is significantly modulated by the emotional properties of the cues used to elicit them. While the dimensional model of emotion typically includes valence (positive vs. negative), arousal (high vs. low intensity), and dominance (controlling vs. controlled), historically (e.g., Talarico et al., 2004), autobiographical memory research has relied on a bi-dimensional model of emotion, focusing on valence (positive vs. negative) and arousal (high vs. low intensity).

Valence is an affective experience that lies on a dimension going from negative to positive, with neutral being in the middle (Russell, 1980). Valence is the cue characteristic inherent to the AMT design, one of the characteristics describing the cues. The Circumplex Model of emotion distinguishes between the valence and arousal characteristics of stimuli, suggesting that while they may represent different emotional experiences, it is vital to consider them simultaneously (Feldman Barrett & Russell, 1998; Russell, 1980). Cue valence appears to function primarily as a determinant of memory accessibility and search strategy (Simpson & Sheldon, 2020). Specifically, positive cues lead to faster reaction times for retrieving memories compared to negative cues suggesting that positive cues facilitate direct access to memory structures, whereas negative cues often require more effortful, generative search strategies to locate specific events.

While researchers traditionally treat valence as a binary category (positive vs. negative) in AMT, this categorical approach may mask significant variance in retrieval mechanisms. Even within the 'positive' category, cues can differ in their emotional density. According to the Self-Memory System (SMS) framework (Conway & Pleydell-Pearce, 2000), cues that are 'highly positive' (e.g., *Successful*) versus 'mildly positive' (e.g., *Interested*) may impose different levels of cognitive demand during the generative search phase. Intensely positive cues may facilitate 'direct access' by bypassing hierarchical search. Conversely, less intense cues may require a 'generative search,' which can increase OGM rates in vulnerable individuals. Thus, treating all positive words as a single block may reduce the AMT's sensitivity to detect subtle differences in memory specificity.

Valence also interacts with the structural properties of cues, such as concreteness, when it comes to memory specificity (Kim et al., 2025). While concrete cues generally elicit specific memories regardless of emotional content, valence becomes critical when cues are abstract. For abstract cues, emotional valence significantly improves the likelihood of retrieving specific rather than overgeneral memories. Furthermore, mood-congruent memory research indicates that the match between cue valence and the memory itself facilitates retrieval, though this effect can be asymmetrical depending on affect regulation strategies (Faul & LaBar, 2023).

While valence directs the search for a specific memory, arousal, the affective experience that differentiates between high and low levels of excitement (Russell, 1980), appears to drive the elaboration and consolidation of that memory (Simpson & Sheldon, 2020). Once an event is accessed, high-arousal cues lead to the reconstruction of more detailed memories compared to low-arousal cues, regardless of the cue's valence. However, the enhancement of detail by arousal is nuanced. While high-arousal events are subjectively rated as more vivid, arousal may selectively enhance "emotion/thought" details rather than perceptual details, particularly for recent memories (Wardell et al., 2021), prioritizing high-salience information for storage at the expense of peripheral details (Faul & LaBar, 2023; Faul et al., 2025). In the AMT, this suggests that high-arousal cues should theoretically decrease OGM by making specific, high-priority memory traces more salient than categorical summaries.

On the other hand, in the context of OGM, excessively high arousal, especially when coupled with negative valence, may trigger functional avoidance mechanisms (see CaR-FA-X model; Williams et al., 2007). In this case, instead of enhancing detail, high arousal may lead the search process to abort prematurely at a categorical level to avoid the emotional distress associated with specific details. Therefore, evaluating arousal ratings is not merely about memory richness, but about understanding the 'bottleneck' where a specific search turns into an overgeneral summary.

Despite the consistent findings regarding valence and arousal, recent critiques argue that these bi-dimensional models blur the specific effects of emotion on memory by ignoring other critical components of emotional experience (Luminet, 2022). A survey of the literature reveals that the majority of studies since 2001 have limited their assessment to arousal and valence. Furthermore, specific discrete emotions within the same valence category (e.g., anger vs. fear) may have distinct effects on memory consistency and vividness due to differences in these underlying appraisals and action tendencies. Therefore, as argued by Luminet (2022), a multi-component approach that includes appraisals and discrete emotional states is recommended to improve the validity of future research to fully capture the complex relationship between emotion and autobiographical memory.

Simpson and Sheldon (2020) investigated how emotional cues that varied in valence and arousal influenced how detailed the events were at retrieval.[2]There is evidence that both arousal and valence of emotional cues contribute to memory retrieval differently. For instance, valence relates to accessibility, whereas arousal relates to increased episodic specificity (Simpson & Sheldon, 2020). Furthermore, the authors confirmed that there is more to emotional cues and their impact on autobiographical memory than only valence. Therefore, the three-factor evaluation of an emotional stimulus, in our case, an emotional word, is worth considering. Especially when emotional cues are known to influence memory properties.

[2] Note that memory details and specificity are not necessarily the same characteristic but could be informative of each other.

As discussed above, the retrieval of autobiographical memories is significantly modulated by the valence and arousal properties of the cues used to elicit them. While the dimensional model of emotion proposed dominance as another critical characteristic in which the human brain evaluates perceptual and symbolic stimuli and represents the sense of control in response to a given stimulus (Bradley & Lang, 1994). Moreover, Verma and Tiwary (2017) showed that using a three-dimensional model involving valence, arousal, and dominance better represents different emotions. Accordingly, Holland and Kensinger (2010) advised accounting for more than one emotional characteristic to appreciate the complexity of emotions and memory. While valence relates to memory accessibility and arousal to episodic detail, dominance may serve as the primary regulator of the generative search phase. Within the framework of the CaR-FA-X model, cues low in dominance evoking feelings of powerlessness may trigger functional avoidance more readily than cues high in dominance. Therefore, examining dominance alongside valence and arousal is not merely an additive step, but a necessary one to understand the stop-rules of autobiographical retrieval. Therefore, simultaneously examining the three cue characteristics would allow us to better comprehend the AMT properties, structure, and applicability.

Dominance represents the degree to which a stimulus elicits a sense of being in control versus being overwhelmed by it (Bradley & Lang, 1994). We suggest that cues characterized by low dominance (e.g., Helpless, Needy) may activate a cognitive schema of powerlessness and for participants with higher depressive symptomatology, these low-dominance cues may act as a signal to stop accessing a specific memory. Recalling a specific event where one felt a total lack of control can be psychologically distressing; thus, the cognitive system may terminate the search at a categorical, overgeneral level as a form of functional avoidance (Williams et al., 2007). This suggestion is consistent with the SMS framework, which suggests that retrieval is a goal-directed process where the working self-monitors the progress of memory search to protect the individual from potentially destabilizing information (Conway & Pleydell-Pearce, 2000). Taken together with findings regarding the effects of valence and arousal, one can suggest that while valence directs the search and arousal provides the energy for detail, dominance appears to set the boundary for how deep into a specific, potentially vulnerable experience the search process is permitted to go. However, since cues used in AMT were selected based on valence and arousal characteristics, the existing literature does not offer data regarding the effects of cue dominance leaving this suggestion largely speculative.

In addition to cue characteristics that potentially influence AMT performance, individual differences are also important to consider because different individuals may appraise the cues differently (Hamann & Canl, 2004). For instance, depressive symptomatology can influence the number of positive events retrieved (Walker et al., 2003) or how positive and negative intensity ratings change over time (Walker et al., 2003). In other words, depressive symptomatology potentially contributes to interpreting emotional cues differently. At the same time, we know from previous research that AMT results depend on depressive symptomatology (Hallford et al., 2021). Hence, since depressive symptomatology can also influence how cues are perceived, it can serve as a plausible and underlying explanation for the differences in AMT results. By investigating how depressive symptoms modulate the appraisal of these three dimensions, this study aims to bridge the gap between stimulus characteristics and individual differences, providing a more focused account of how cue perception influences memory specificity.

2. The Present Study

The present study aims to apply a three-factor emotional stimuli evaluation strategy and investigate how emotional cues are interpreted based on valence, arousal, and dominance of the words and whether these interpretations change due to participants' depression levels. In addition, we explored whether different levels of depressive symptomatology are related to how emotional words are appraised and whether other emotional characteristics provide more information besides valence.

Our hypotheses are listed below. Since the study has a explorative nature, the hypotheses stated are non-directional.

1. In addition to being more positive, positive cue words will differ from negative cue words for arousal and dominance as well.

2. We expect the cue words in the same category to differ from each other for valence, arousal and dominance.

3. We also expect these three cue words characteristics to be related to depression levels.

3. Method

3.1. Participants

The study sample consisted of 40 University Students (20 female) with a mean age of 22.25 (*SD* = 3.90) whose native language was Turkish. For our main analysis, one-way within-subjects ANOVA with five measures, the G*Power3 analysis revealed a need for 31 participants for $p = .05$, power =.95 (Faul et al., 2007). Mean BDI-II scores of the participants were 9.14 (*SD*= 4.87) with a maximum of 19 after removing 4 outliers. As compensation, the participants received course credit. The sample was non-clinical as the participants who reported being diagnosed with a mental health illness were excluded from the analysis. Participants who were diagnosed with a mental disorder were excluded from all analyses. In addition, none of the participants reported using medication that may affect cognitive processes.

3.2. Materials

Cue Words. Studies published between 1986 and 2015 were screened to determine the cue words used in OGM studies. We examined 34 published studies between 1986 and 2015 (the year of data collection). Amongst these 34 studies, only 21 of them reported the cue words they used. Sixty-eight words were selected according to their translatability into Turkish as one word. For example, the word *happy* can be translated into Turkish as one word, "mutlu" but the word *acceptable* can only be translated as two words, "kabul edilebilir". The whole list of words, along with our participants' ratings of valence, arousal, and dominance can be seen on our OSF page. Depending on the percentage of usage in the OGM studies, we selected a list of 5 positive [happy (.71), surprised (.52), successful (.57), safe (.52), interested(.52)] and 5 negative [angry (.71), lonely (.67), hurt (.62), clumsy (.48), sorry (.33)] most commonly used cue words for further analyses due to their frequent usage and space limitations.

Self-Assessment Manikin (SAM; Bradley & Lang, 1994). SAM is a non-verbal pictorial tool used to measure different affective dimensions related to different types of stimuli. The present study used five-point Likert SAMs to measure subjective valence, arousal, and dominance ratings related to different words. SAM ranges from a smiling figure (positive affect) to a frowning figure (negative affect) for valence, from an excited, wide-eyed figure (high arousal) to a relaxed, sleepy figure (low arousal) for arousal and from a small figure (feeling controlled) to a large figure (feeling in-control) for dominance.

Beck Depression Inventory (BDI-II; Beck et al., 1996). The BDI-II measures the severity of depression symptoms in the last 2 weeks. It consists of 21-self report items rated on a scale ranging from 0 (low symptomology) to 3 (high symptomology). Kapci et al. (2008) confirmed the Turkish version to be valid and reliable, with a Cronbach's alpha of .90.

3.3. Procedure

Data were collected online using Google Forms. After participants were shown informed consent and agreed to participate in the study, the participants were asked to pick the corresponding figure for how the cue word presented made them feel. The cue words were presented in a fixed random order. Specifically, a single randomized sequence was generated and then presented consistently to all participants to ensure that any potential order effects remained constant across the sample. Participants rated each word for valence, arousal and dominance in a fixed order using separate SAMs for each dimension. After the cue word ratings were completed, participants were asked to complete the BDI-II, and demographic questions. The procedure lasted 20 minutes on average. The data collection took place in two waves in 2015.

3.4. Analysis Plan

Means and standard deviations were calculated for the 68 cue words (see the OSF page above for the complete list). As mentioned in the Method section, ratings of 5 most commonly used positive and 5 negative cue words were analyzed further. To facilitate the selection of standardized cue sets, we established threshold-based classifications for valence, arousal, and dominance. These thresholds were determined based on the 5-point Likert scale used in the SAM, where scores above 3.5 were categorized as 'high' (representing positive affect, high excitement, or high control) and scores below 2.5 were categorized as 'low' (representing negative affect, low excitement, or low control). This allows for the identification of words with the most distinct affective profiles. While the current data structure involves multiple ratings nested within both participants and items, we employed repeated-measures ANOVA and t-tests as a primary analytical approach to identify broad differences in cue characteristics. We recognize that these traditional methods treat item-level variability as a fixed effect rather than a random effect. However, given the exploratory nature of this study and the specific focus on within-subject appraisal of the ten most common AMT cues, these methods were deemed appropriate for initial hypothesis testing. First, the mean ratings of valence, arousal, and dominance for the 5 positive and 5 negative cue words were compared with each other using within-subjects t-tests. Furthermore, we used separate one-way repeated measures ANOVAs to compare the specific positive cue words within each other with regards to the three stimuli evaluation categories (valence, arousal, dominance) and negative cue words within each other, also in terms of valence, arousal, and dominance. The pairwise comparisons were conducted using Bonferroni corrections. Pearson's correlation test was used to explore the relationship between BDI-II scores and cue word ratings. The two-tailed significance values were reported.

4. Results

4.1. Cue Word Comparisons

4.1.1. Comparisons of Positive and Negative Cue Words

The mean valence, arousal, and dominance ratings given to positive and negative cue words were compared. The mean valence ratings of the five most commonly used positive (M = 4.23, SD = 0.35) and five most commonly used negative (M = 1.55, SD = 0.41) words were significantly different from each other, [$t(39) = 27.30$, $p < .001$], showing that positive words were indeed rated as more positive than the negative ones. Mean arousal ratings of positive (M

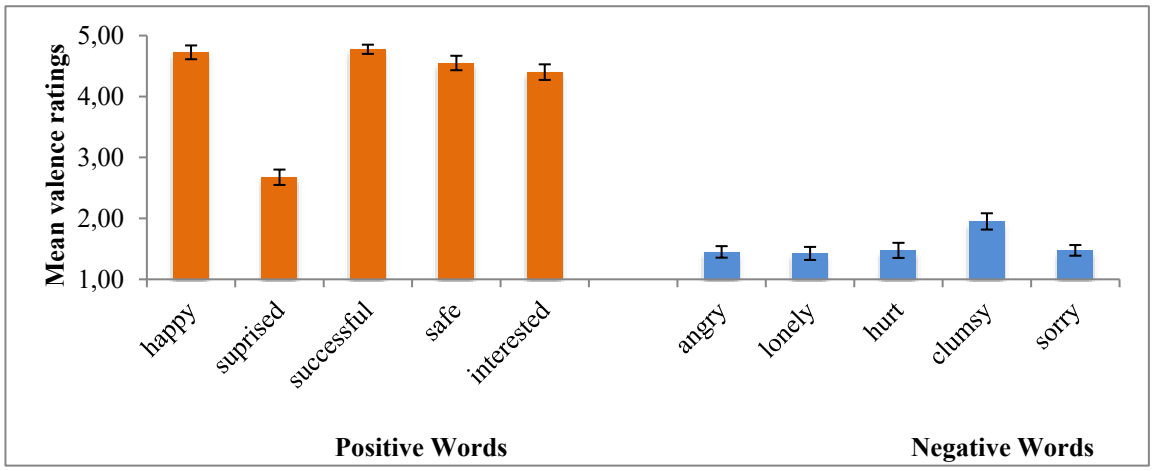
= 3.35, *SD* = 0.62) and negative (*M* = 3.18, *SD* = 0.69) words were not different from each other (*p* = .265). Finally, for dominance ratings, positive cue words were rated as more dominant (*M*= 3.79, *SD* = 0.48) than negative cue words (*M* = 2.89, *SD* = 0.73) words *t*(39)= 6.80, *p* < .001, suggesting that valence is not the only dimension differentiating the positive and negative cue words from each other. These results partially support hypothesis 1.

4.1.2. Comparisons of Valence Ratings

When the valence ratings of positive cue words were explored, the results revealed a main effect of a specific cue word on the rated valence scores of the words, *F*(2.62, 102.51)= 62.50, *p* < .001, *MSE* = 47.19, n_p^2 = .616. The large effect size suggests that the specific cue word choice is a primary determinant of the variance observed in valence ratings. Pairwise comparisons revealed that the cue word *surprised* (*M* = 2.68, *SD* = 0.80) was rated as less positive than *happy* (*M* = 4.72, *SD* = 0.72), *successful* (*M* = 4.77, *SD* = 0.48), *safe* (*M* = 4.55, *SD* = 0.75), and *interested* (*M* = 4.40, *SD* = 0.81) (all *p*s < .001), (see Figure 1 for a graphical depiction of results). Furthermore, the cue word *interested* was rated as less positive than *happy* (*p* = .019). While other differences between positive words were not significant, the results suggest that the 5 most commonly used positive words were not equally positive.

For negative words, the main effect of cue word on valence was significant, *F*(2.97, 115.88)= 4.88, *p* = .003, *MSE* = 2.65, n_p^2 = .111. According to the pairwise comparisons, *clumsy* (*M* = 1.95, *SD* = 0.85) was rated as less negative than the words *angry* (*M* = 1.45, *SD* = 0.60), *lonely* (*M* = 1.42, *SD* = 0.68), and *sorry* (*M* = 1.48, *SD* = 0.55), all *p*s < .05 (see Figure 1). All other comparisons were non-significant (all *p*s > .05). These results showed that not all negative cue words were equally negative.

Figure 1: Means of valence ratings for 5 most commonly used positive and 5 negative cue words



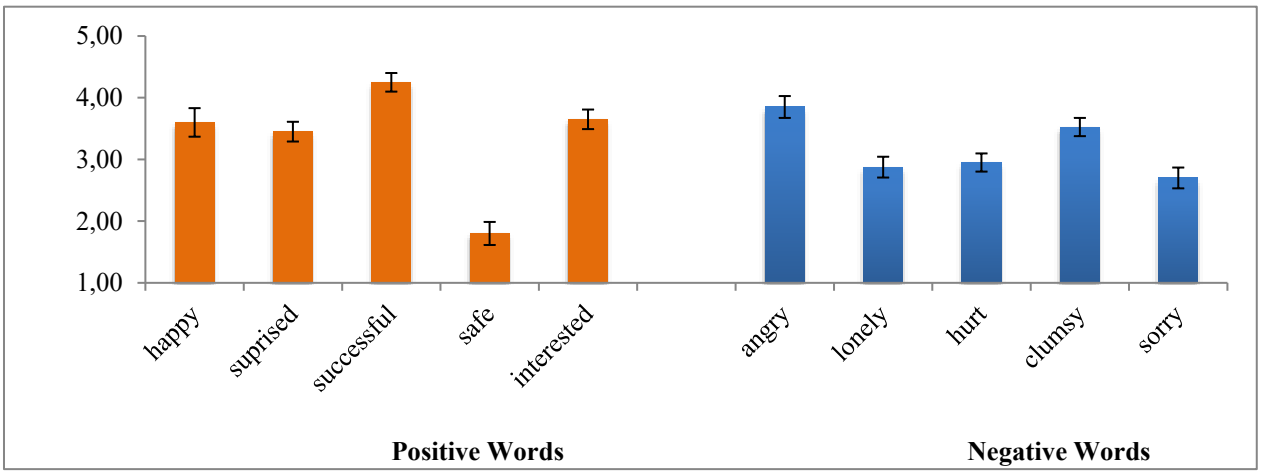
4.1.3. Comparisons of Arousal Ratings

When positive words were compared amongst themselves, a main effect of cue words on arousal ratings was found, *F*(4, 156)= 13.17, *p* < .001, *MSE* = 9.44, n_p^2 = .252. The large effect size suggests robust differences in arousal levels even within the same valence category. Pairwise comparisons showed that for arousal, the cue word *safe* (*M* = 1.80, *SD* = 0.60) received lower ratings than cue words *successful* (*M* = 4.25 *SD* = 0.95), *happy* (*M* = 3.60, *SD* = 1.46), *surprised* (*M* = 3.45, *SD* = 1.01), and *interested* (*M* = 3.65, *SD* = 1.00), all *p*s < .05. Figure 2 present a visual depiction of the results. Furthermore, the cue word *successful* was rated as higher on arousal than the words *surprised* (*p* < .001) and *interested* (*p* = .045). Other comparisons were not significant. The results suggest that in addition to their differences in

valence, the most commonly used 5 positive words are also different from each other in terms of arousal.

For the arousal ratings of the negative words, there was a main effect of a cue word as well, $F(4, 156)= 13.17$, $p < .001$, $MSE = 9.44$, $n_p^2 = .252$. Pairwise comparisons revealed that *angry* ($M = 3.85$, $SD = 1.12$), and *clumsy* ($M = 3.53$, $SD = 0.93$), received higher arousal ratings than *lonely* ($M = 2.87$, $SD = 1.06$), *hurt* ($M = 2.95$, $SD = .93$), and *sorry* ($M = 2.70$, $SD = 1.07$), all *p*s < .05. No other comparison revealed significant results.

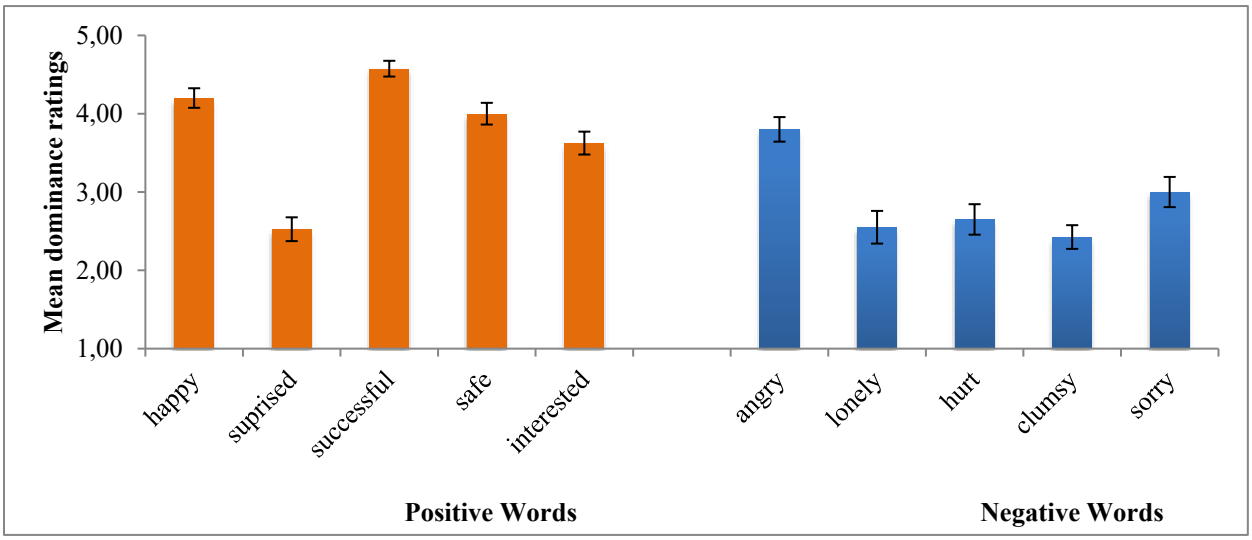
Figure 2: Means of arousal ratings for 5 most commonly used positive and negative cue word



4.1.4. Comparisons of Dominance Ratings

Similar to valence and arousal, cue words had a main effect on dominance ratings, $F(4, 156)= 43.12$, $p < .001$, $MSE = 26.88$, $n_p^2 = .525$. The word *surprised* ($M= 2.52$, $SD = .96$) received the lowest dominance ratings among the 5 positive words (see Figure 3), all *p*s < .001. In addition to the cue word *surprised*, the word *successful* ($M= 4.57$, $SD = 0.63$) was rated as more dominant than the words *safe* ($M= 4.00$, $SD = .88$; $p = .003$) and *interested* ($M= 3.62$, $SD = 0.93$; $p < .001$). Finally, the word *happy* ($M= 4.20$, $SD = .79$) was rated more dominant than the word *interested* ($M= 3.62$, $SD = .93$), $p = .009$.

Figure 3: *Means of dominance ratings for 5 most commonly used positive and negative cue words*



There was a main effect of cue words on dominance for the negative words, $F(4, 156)= 12.39$, $p < .001$, $MSE =12.30$, $n_p^2 = .241$. *Angry* ($M= 3.80$, $SD = 0.99$) was rated as more dominant than *lonely* ($M= 2.55$, $SD = 1.31$), *hurt* ($M= 2.65$, $SD = 1.23$), *clumsy* ($M= 2.42$, $SD = 0.95$) and *sorry* ($M= 3.00$, $SD = 1.22$), all *p*s <. 01. The remaining comparisons were not significant. Together, these results support hypothesis 2 and draw attention to differences in the cue word characteristics.

Figure 4: Distribution of 5 most commonly used positive and negative cue words according to valence, arousal, and dominance

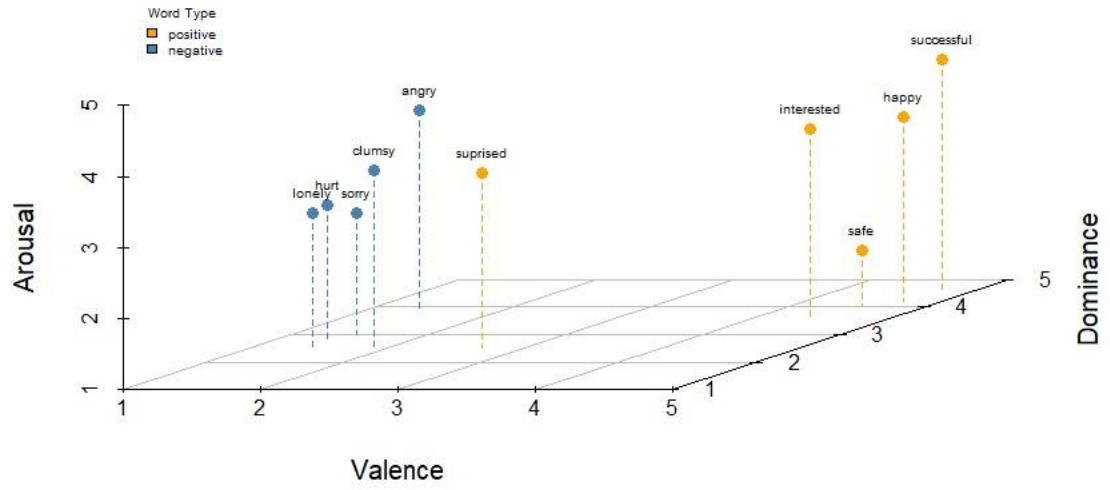


Table 1: Distribution of the positive (> 3.5) and negative (< 2.5) cue words according to high (> 3.5) and low (< 2.5) arousal and dominance ratings.

		Dominance	
Valence	Arousal	High	Low
Positive	high	Able, Ambitious, Brave, Courage, Efficient, Energetic, Enthusiastic, Friendly, Funny, Happy, Interested, Pleasurable, Social, Successful	-
	low	Honest, Peaceful, Relieved, Safe, Trust	-
Negative	high	Afraid, Angry, Guilty, Hope, Tense	Clumsy
	low	-	Inefficient, Lazy, Useless

4.2. Cue Words Ratings and Depression Levels

Pearson's Correlation Analyses were conducted to see whether valence, arousal, and dominance ratings of different cue words relate to BDI-II scores differently. After conducting the

correlation analyses, we counted the cue words that are significantly related to depression and calculated their percentage among all cue words. For valence, 6% of the cue words were significantly related to depression levels, for arousal, 10% were related and finally, for dominance, 6% of the cue words were related to BDI-II, 3. The results of the analyses for all of the 68 cue words can be seen at OSF Repository.

Regarding the most commonly used 5 positive words, for the cue word *happy*, results revealed a negative correlation between its valence ratings and BDI-II scores ($r(38) = -.346$, $p = .029$);the higher the positive rating of *happy*, the lower were BDI-II scores. On the other hand, arousal ratings of the cue word *surprise*, a cue word commonly used as a positive word, were positively correlated with BDI-II ($r(38) = .334$, $p = .035$) indicating that if the word *surprise* is considered higher in arousal, it is associated with higher scores of depressive symptoms. For the 5 positive cue words, none of the other correlations were significant. For the most commonly used five negative cue words none of the correlations were significant.

Table 2: Correlations between the valence, arousal and dominance ratings of 5 positive and 5 negative words and BDI-II scores.

	Valence		Arousal		Dominance	
	r	p	r	p	r	p
happy	-.346*	.029	.025	.876	.024	.883
surprised	.075	.646	.334*	.035	-.016	.920
successful	.185	.253	-.013	.936	-.144	.374
safe	.159	.328	-.094	.565	.006	.969
interested	-.283	.077	-.294	.065	-.109	.503
angry	.025	.880	-.005	.975	.172	.289
lonely	-.165	.308	.071	.663	.132	.416
hurt	-.250	.120	-.136	.401	.204	.207
clumsy	-.045	.782	-.044	.789	.073	.656
sorry	-.223	.167	.112	.493	.105	.519

5. Discussion

In the current study, we examined whether emotional cue words differed regarding valence, arousal, and dominance and whether the ratings of Turkish emotional words were associated with individual characteristics of depression in a non-clinical sample. The original AMT developed by Williams and Broadbent (1986) consisted of 5 positive and 5 negative emotional cues. Since 1986 the design of AMT has changed substantially. However, the cues themselves have rarely been evaluated. This study points to the importance of acknowledging the differences between cues. Based on emotional stimuli appraisal literature (Russell, 1980) and our findings, we argue that the words differ substantially regarding their valence, arousal, and dominance. The high effect sizes observed across valence, arousal, and dominance dimensions further emphasize that these differences are not merely statistically significant but practically meaningful for experimental design. The variance across these three characteristics may help clarify some of the contradictory findings reported in the literature regarding positive and negative cues (Hallford et al., 2021). Moreover, when we looked at the cues within each category (i.e., positive and negative), we saw that they do not necessarily form one entity and may differ. Further, we found that the ratings of the cues are related to depression scores.

Furthermore, we found that individual differences in depression were associated with the ratings of the emotional words. Higher depression scores were associated with lower positive ratings of *happ*iness. This finding implies that the positive cue words presented to depressed participants may be more negative for them. Most OGM studies (e.g., Dritschel et al., 2014), compare and subsequently combine specificity scores for positive and negative cues. This common practice stems from a lack of significant differences between valence categories, often attributed to similar underlying processing. However, positive and negative cue categories showed significantly different valence ratings but were similar in arousal. The fact that arousal did not significantly differ between positive and negative cue words may explain the reported null effects between positive and negative words. When considered together with the findings of a reduced positive affect in people with depression (Anderson et al., 2023), this implies that depression may change how people perceive positive words. This would suggest that the analysis of positive and negative cue categories needs to be re-considered in more detail because of the differences in ratings of specific cues. For instance, for the word *surprise*, depression was associated with higher arousal ratings. The arousal of the positive word *surprise* could also imply certain qualitative meaning it might have for people with depression, for whom uncertain (i.e., surprising) situations are instead threatening (Beck et al., 1987).

It is important to note that our study covered a non-clinical sample which is relevant because many studies considering the relationship between OGM, AMT, and depression usually cover clinical ones (Hallford et al., 2021). Thus, if even in a non-clinical sample, depression and anxiety scores impact how the cue words are perceived, this effect could even be more substantial in clinical samples.

Although the present study did not directly assess AMT performance, variability in the emotional properties of cue words may have implications for how AMT findings may be interpreted. In AMT research, cue words are typically grouped into broad positive and negative categories. However, our results suggest that cues within the same valence category differ meaningfully, which could potentially lead to the engagement of different memory retrieval processes.

The ten most commonly used words in OGM research showed significant differences between positive and negative cues in terms of valence. Positive cue words were rated as more positive than negative cue words. This aligns with the vast amount of research that differentiates between positive and negative categories in the AMT (e.g., Roberts et al., 2018; Wessel et al., 2014). In the present dataset, no systematic differentiation in arousal was observed at the level of broad valence categories. Rather than indicating an absence of arousal-related differences between positive and negative stimuli in general, this pattern suggests that arousal may vary primarily at the level of individual cue words within each category. We found significant differences in the dominance ratings of positive and negative words, showing higher dominance ratings for positive words. While traditional research focuses on binary valence and arousal, our results identify Dominance (Agency) as a critical third dimension. This dimension significantly reorganizes the psychological profiles of emotional cues. By applying thresholds (means > 3.5 or < 2.5) across all three dimensions, we identified a subset of cue words characterized by highly differentiated affective profiles (see Table 1). As shown in Table 1, this classification system highlights word clusters that researchers can use to match stimuli. For example, while ‘Happy’ and ‘Successful’ both meet the criteria for high-valence and high-dominance, they differ in their specific intensities. This mapping provides a practical guide for selecting alternate cues that maintain specific emotional properties while varying others, such as matching negative words for arousal while differing in perceived dominance. The relationship between dominance and OGM, however, needs further research. Current results

highlight that each cue can be perceived differently, even within the same valence category. Thus, examining additional features of cue words other than valence is essential.

Previous studies usually did not investigate the valence ratings and the differences among cues within the categories (e.g., Ricarte et al., 2014; Sumner et al., 2014). The present study compared the most commonly used positive and negative cue words within categories. We found that specific cues were evaluated as more positive or negative than the others. For instance, the word *surprised* was rated as the least positive compared to other positive cues, or *clumsy* was less negative than *angry*, *lonely*, and *sorry*. Variation in valence within positive or negative cue sets may be relevant for the initial accessibility of autobiographical memories. According to the Self-Memory System framework (Conway & Pleydell-Pearce, 2000), some cues may facilitate relatively direct access to specific memories, whereas others require a more effortful, generative search. Cues that are only weakly positive or affectively ambiguous may therefore be less effective in eliciting specific memories than more strongly valenced cues, even when categorized under the same label.

Our results also showed that, similarly to the valence ratings, the words differed within each category for arousal and dominance ratings. Notably, *safe* scored the lowest in arousal among the positive words, while *angry* and *clumsy* were the highest for the negative words. Although arousal did not differentiate positive and negative categories at an aggregate level, substantial variability was observed across individual cue words. Given that arousal is linked to attentional prioritization and mnemonic processing (Mather & Sutherland, 2011), such variability may influence the vividness or specificity of retrieved memories, but would be obscured when cues are selected solely on the basis of valence.

In addition, *surpris*ing was the least dominant in the positive categories, whereas *ang*er was the most dominant among the negative words. Dominance may be particularly relevant in the context of AMT, as cues associated with low perceived control may activate themes of helplessness or reduced agency, potentially encouraging responses at a more general level. While this possibility remains speculative, the availability of dominance ratings allows future studies to examine whether agency-related cue properties contribute to variability in AMT performance beyond valence and arousal. It is critical to distinguish these theoretical possibilities from our empirical findings; as our study did not directly measure memory specificity, the proposed role of dominance in setting the 'boundary' for memory search remains a hypothesis requiring further empirical testing via the AMT. Taken together, these considerations highlight the value of accounting for the multidimensional emotional properties of cue words when designing and interpreting AMT studies. Our findings support the assumption that the prominent categorization into positive and negative valence cues may benefit from closer consideration of additional affective dimensions.

5.1. Alternative Cue Words for Memory Research

The present study provides a preliminary set of multidimensional ratings for emotional cue words in a Turkish-speaking sample. These preliminary ratings provide a basis for the identification of alternative cue word sets matched on valence and arousal but differing in dominance, thereby offering researchers finer experimental control. For instance, within the negative, high-arousal words, words such as Afraid (Korkmuş), Angry (Kızgın), Tense (Gergin), and Worried (Endişeli) exhibit comparable arousal levels but vary in dominance, making them suitable alternatives to canonical fear-related cues. Similarly, in the positive, low-arousal words, Peaceful (Huzurlu), Relieved (Rahatlamış), and Relaxed (Rahat) provide alternatives to commonly used calmness-related words. In the negative, low-arousal range, cues

such as Guilty (Suçlu), Regretful (Pişman), and Disappointed (Hayal kırıklığına uğramış) may serve as substitutes for sadness-related cues.

6. Limitations and Future Directions

The current study has several limitations. First of all, OGM is mainly observed in people diagnosed with depression. However, in the current study, we used an analog sample. Future studies should investigate the ratings of cue words in clinical samples to see whether the present results can be generalized. In this study, the emotional words were translated into the Turkish language. Languages may present different variations in word translations and perception (Kapucu et al., 2021). Thus, comparing the current ratings with ratings of the words presented in other languages could be beneficial and contribute to a more standardized method of the AMT design.

Furthermore, while translatability was a primary practical criterion for ensuring linguistic consistency across the Turkish and English versions of the AMT, we recognize that this approach does not account for other psycholinguistic properties known to affect memory retrieval. However, other psycholinguistic variables such as word frequency, familiarity, imageability, and concreteness were not formally controlled in our cue selection process. Future research should incorporate these variables into a more comprehensive selection model to further refine the methodological precision of the AMT.

Moreover, a sample covering the entire adult lifespan could provide more information about the differences in age groups regarding evaluating certain stimuli, especially considering the findings regarding the positivity effect phenomenon (Carstensen et al., 1999). Younger adults tend to process negative stimuli more thoroughly, whereas this tendency shifts in favor of positive stimuli as the participants' ages. Furthermore, while the sample size provides sufficient statistical power to detect within-subject differences in cue ratings, we acknowledge that this analogue sample of university students is not intended to represent a broad normative population and these results should be viewed as preliminary. Establishing stable normative values for the Turkish population would require significantly larger and more demographically diverse samples. Therefore, future research should examine whether the ratings of AMT cue words with larger sample sizes across different age groups.

Another methodological limitation of the present study concerns the statistical handling of the nested data structure. Our analyses utilized repeated-measures ANOVA and Pearson correlations, which are standard in OGM research but do not fully account for the simultaneous dependency between participant-level and item-level variability. Future research with larger sample sizes should employ Linear Mixed-Effects Models or Multilevel Modeling. Such approaches would allow for a more nuanced understanding of how specific word-level properties interact with individual participant characteristics (e.g., depression levels) to influence affective appraisals.

Finally, it should be noted that data collection occurred in 2015. While the core emotional cues of the AMT are linguistically stable, the temporal gap should be considered, as cultural connotations and affective associations with certain abstract words may evolve over time.

Conclusion

In this study, we stressed the importance of different emotional word characteristics in addition to positive and negative valence categorization. Our findings show that the cues within the said category can have different ratings of valence, arousal, and dominance. Additionally, to cue characteristics that may be differently associated with different word cues (e.g.,

surprised, happy, angry), depression and anxiety contributed to the ratings of emotional words. This is important to consider in the line of research where various cues are used to elicit different memories. Taken together, the present dataset serves as a pilot framework for more controlled cue selections and a flexible toolbox for exploratory and theory-driven research. We suggest that moving beyond frequency-based cue selection and incorporating dominance alongside valence and arousal will enable researchers to more precisely tailor emotional stimuli to their theoretical questions, thereby increasing the dataset's applied value. Future research should focus on developing more standardized versions of the emotional cue words for memory research.

Information Note

I have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript. I have no conflicts of interest to report.

All data, materials, and additional results are available at https://osf.io/3kcs8/?view_only=83be6923507a4b67817e61602224f963.

All participants in this research received informed consent, were assured confidentiality, and treated according to ethical code of conduct. I have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript. I have no conflicts of interest to report.

I would like to thank Dr. Justina Pociūnaitė for her contributions to the first drafts of this paper.

References

Anderson, R. J., Clayton McClure, J. H., Boland, J., Howe, D., Riggs, K. J., Dewhurst, S. A. (2023). The relationship between depressive symptoms and positive emotional anticipation of goal achievement. *Journal of Experimental Psychopathology*, *14*(1), 20438087231164963.

Bartoli, E. & Smorti, A. (2018). Facing the Language-Memory Problem in the Study of Autobiographical Memory. *Integrative Psychological and Behavioral Science*, *53*(3), 374-396.

Beck, A. T., Kendall, P. C., Hollon, S. D., Hammen, C. L., Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research, 11(*3), 289-299.

Beck, A. T., Steer, R. A., Ball, R., Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment*, *67*(3), 588-597.

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.

Conway, M. A. & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261-288.

Dritschel, B., Beltsos, S., McClintock, S. M. (2014). An "alternating instructions" version of the Autobiographical Memory Test for assessing autobiographical memory specificity in non-clinical populations. *Memory*, *22*(8), 881-889.

Faul, F., Erdfelder, E., Lang, A. G., Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Faul, L. & LaBar, K. S. (2023). Mood-congruent memory revisited. *Psychological Review*, *130*(6), 1421.

Faul, L., Ritchey, M., Kensinger, E. A. (2025). The relationship between subjective vividness and remembered visual characteristics of emotional stimuli across the lifespan. *Emotion*, *25*(6), 1579-1595.

Feldman Barrett, L. & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967-984.

Gollan, J. K., Hoxha, D., Hunnicutt-Ferguson, K., Norris, C. J., Rosebrock, L., Sankin, L., Cacioppo, J. (2016). The negativity bias predicts response rate to Behavioral Activation for depression. *Journal of Behavior Therapy and Experimental Psychiatry*, (52), 171-178.

Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., Zinbarg, R. E., Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory*, *17*(6), 609-623.

Hallford, D., Rusanov, D., Yeow, J., Barry, T. (2021). Overgeneral and specific autobiographical memory predict the course of depression: An updated meta-analysis. *Psychological Medicine, 51*(6), 909-926.

Hamann, S. & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology*, *14*(2), 233-238.

Holland, A. C. & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews*, *7*(1), 88-131.

Kapci, E. G., Uslu, R., Turkcapar, H., Karaoglan, A. (2008). Beck Depression Inventory II: evaluation of the psychometric properties and cut-off points in a Turkish adult population. *Depression and Anxiety*, *25*(10), E104-E110.

Kapucu, A., Amado, S., Eroğlu, S., İyilikçi Arıkan, E. (2018). Kodlama öncesi veya geri getirme öncesi sevk edilen korku ve kızgınlık duygularının tanıma belleği performansına etkileri. *Türk Psikoloji Dergisi, 33*(82), 85-99.

Kim, H., Harris, C. B., Barber, S. J. (2025). Autobiographical memory specificity in younger and older adults as a function of cue type. *Memory*, 1-14.

Luminet, O. (2022). Towards a better integration of emotional factors in autobiographical memory. *Memory*, *30*(1), 49-54.

Mather, M. & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*(2), 114-133.

Ricarte, J. J., Hernández, J. V., Latorre, J. M., Danion, J. M., Berna, F. (2014). Rumination and autobiographical memory impairment in patients with schizophrenia. *Schizophrenia Research*, *160*(1-3), 163-168.

Ricarte, J. J., Latorre, J. M., Ros, L. (2014). Effects of instructions and cue subjectiveness on specificity of autobiographical memory recall. *Anales De Psicología*, *30*(3).

Roberts, J. E., Yanes-Lukin, P., Kyung, Y. (2018). Distinctions between autobiographical memory specificity and detail: Trajectories across cue presentations. *Consciousness and Cognition*, *65*(August), 342-351.

Ros, L., Romero, D., Ricarte, J. J., Serrano, J. P., Nieto, M., Latorre, J. M. (2018). Measurement of overgeneral autobiographical memory: Psychometric properties of the autobiographical memory test in young and older populations. *PLOS ONE*, *13*(4), e0196073.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178.

Simpson, S. & Sheldon, S. (2020). Testing the impact of emotional mood and cue characteristics on detailed autobiographical memory retrieval. *Emotion*, *20*(6), 965-979.

Sumner, J.A. (2012). The mechanisms underlying overgeneral autobiographical memory: An evaluative review of evidence for the CaR-FA-X model. *Clinical Psychology Review, 32*(1), 34-48.

Sumner, J. A., Mineka, S., McAdams, D. P. (2013). Specificity in autobiographical memory narratives correlates with performance on the autobiographical memory test and prospectively predicts depressive symptoms. *Memory, 21*(6), 646-656.

Sumner, J.A., Mineka, S., Zinbarg, R. E., Craske, M. G., Vrshek-Schallhorn, S., Epstein, A. (2014). Examining the long-term stability of overgeneral autobiographical memory. *Memory*, *22*(3), 163-170.

Talarico, J. M., LaBar, K. S., Rubin, D. C. (2004). Emotional intensity predicts autobiographical memory experience. *Memory & Cognition*, *32*(7), 1118-1132.

Verma, G. K. & Tiwary, U. S. (2017). Affect representation and recognition in 3D continuous valence-arousal-dominance space. *Multimedia Tools and Applications*, *76*(2), 2159-2183.

Walker, W. R., Skowronski, J. J., Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way!. *Review of General Psychology*, *7*(2), 203-210.

Wardell, V., Madan, C. R., Jameson, T. J., Cocquyt, C. M., Checknita, K., Liu, H., Palombo, D. J. (2021). How emotion influences the details recalled in autobiographical memory. *Applied Cognitive Psychology*, *35*(6), 1454-1465.

Wessel, I., Postma, I. R., Huntjens, R. J. C., Crane, C., Smets, J., Zeeman, G. G., Barnhofer, T. (2014). Differential correlates of autobiographical memory specificity to affective and self-discrepant cues. *Memory*, *22*(6), 655-668.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*(1), 122-148.

Williams, J. M. G. & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology*, *95*(2), 144-149.

Williams, J. M. G., Healy, H. G., Ellis, N. C. (1999). The effect of imageability and predicability of cues in autobiographical memory. *The Quarterly Journal of Experimental Psychology Section A, 52*(3), 555-579.

Williams, J. M. G. & Scott, J. (1988). Autobiographical memory in depression. *Psychological Medicine*, *18*(3), 689-695.