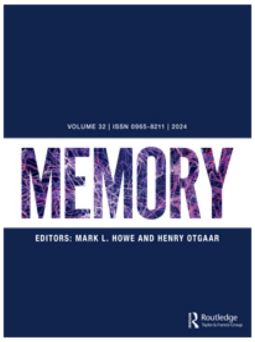

Memory

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pmem20

The effect of emotional valence and font size on metacognition and memory

Karina Agadzhanyan & Alan D. Castel

MEMORY
2024, VOL. 32, NO. 2, 252–263
https://doi.org/10.1080/09658211.2024.2307919

Routledge
Taylor & Francis Group

The effect of emotional valence and font size on metacognition and memory*

Karina Agadzhanyan and Alan D. Castel

Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA

ABSTRACT
Predictions about memory involve the use of metacognition, and metacognition can rely on various cues. The present study investigated metacognition and recall performance when to-be-remembered words differed in font size and emotional valence, to determine what cues are utilised when making metacognitive judgments. Participants were presented with lists of words varying in font size (small and large) and emotional valence (negative and neutral) and were asked to remember as many words as possible for a later recall test while engaging in item-level metacognitive assessments. Specifically, after studying each word, participants either made only judgments of learning (JOLs, Experiment 1) or both JOLs and restudy judgments (Experiment 2). Across experiments, results revealed that while JOLs were sensitive to both font size and emotional valence, restudy judgments were mostly sensitive to emotional valence, and participants' metacognitive assessments mapped onto memory performance generally for emotional words. Additionally, we found that the effect of font size on metacognition and memory was robust to experience-based learning. Together, the current study extends our understanding of how emotion and font size affect metacognition (monitoring and control) and memory and suggests that when presented with multiple cues, certain diagnostic cues can be harnessed to mitigate metacognitive illusions.

ARTICLE HISTORY
Received 18 May 2023
Accepted 10 January 2024

KEYWORDS
Memory; metacognition; emotion; font size; judgments of learning

In our daily lives, we are often exposed to an enormous amount of information, whether it's on paper, displayed on screens, or from various other sources in our environment. Generally, emotional information is remembered better than non-emotional (i.e., neutral) information, known as the *emotion saliency effect*, and this tendency has been observed in laboratory settings with both pictures and words (see Kensinger, 2009; Murphy & Isaacowitz, 2008, for reviews). When studying words for a memory test, emotional valence, among other stimuli characteristics, like animacy, concreteness, frequency, and word length, is known to affect memorability (Kensinger & Corkin, 2003; Murphy & Castel, 2022). However, emotional information may affect memory differently when combined with other cues available at encoding, such as intensity, font size, volume, or frequency.

Researchers often incorporate metacognitive assessments to investigate how stimuli characteristics differentially impact one's predictions of future memory performance and actual memory performance. Metamemory, or one's own awareness and knowledge of their memory abilities, consists of both metacognitive monitoring or self-assessment of one's learning and memory, and metacognitive control, which refers to the self-regulation of learning based on monitoring (Dunlosky et al., 2016; Dunlosky & Metcalfe, 2008; Nelson & Narens, 1990; Rivers et al., 2020). To investigate metacognitive monitoring, researchers often use *judgments of learning* (JOLs), which are prospective memory assessments or predictions of the likelihood of remembering a given item on a future test (see Rhodes, 2016 for a review). Studies examining the effect of emotion on metacognitive monitoring have found that emotionality has an enhancing effect on one's predictions of future recall, as often measured through their JOLs. Termed the *emotion saliency effect on JOLs* (Tauber et al., 2017; see Yin et al., 2023 for a review), it has been shown that people perceive emotional words as more likely to be remembered, thus giving higher JOLs to emotional than neutral information (Hourihan et al., 2017; Nelson & Narens, 1990; Zimmerman & Kelley, 2010). Therefore, emotional valence not only impacts metamemory but is also diagnostic of future memory performance.

It has also been shown that metacognitive monitoring informs control processes, such that learners use their JOLs as a basis for regulating the allocation of study time for to-be-learned information (Nelson, 1996). For example, according to the discrepancy reduction theory, when encountered with unfamiliar or new information, learners seek to reduce the discrepancy between their

CONTACT Karina Agadzhanyan karadefrost@g.ucla.edu Department of Psychology, University of California, Los Angeles, CA, 90095, USA
*The experiments reported in this article were formally preregistered. The stimuli and data have been made available on the Open Science Framework https://osf.io/9g4nb/.

current and optimal state of learning (Dunlosky & Hertzog, 1998; Thiede & Dunlosky, 1999). Specifically, learners are expected to rely on metacognitive monitoring processes by selecting to study items they perceive as more difficult to learn as opposed to items they perceive as easier. By doing so, they aim to narrow the discrepancy between their current knowledge level and desired level of learning (Dunlosky & Thiede, 2004).

Furthermore, metacognitive judgments are often informed by the features or cues available at the time of the encoding. According to Koriat's (1997) cue-utilisation theory, there are three types of cues that influence metacognition: extrinsic cues, intrinsic cues, and mnemonic cues. While variables that affect an item's characteristics without altering its objective nature, such as presentation time, number of study presentations, or perceptual processing fluency (e.g., font size), belong to the category of *extrinsic cues*, factors that inform an item's internal nature, such as word frequency, concreteness, or emotionality belong to the category of *intrinsic cues.* And lastly, *mnemonic cues*, refer to one's prior experience, such as the ease with which an item comes to mind. Based on Koriat's theory (1997), individuals are more likely to base their JOLs on intrinsic cues, such as their understanding of the material, rather than extrinsic cues, such as study conditions. For example, when predicting one's own memory performance, individuals tend to rely more on emotional valence or word frequency (intrinsic cues) than on font size or presentation time (extrinsic cues). Additionally, exposing participants to multiple study-test cycles revealed that as participants' learning increased, they tended to become more underconfident in their JOLs, indicating that their estimates of how well they had learned were lower than what their actual performance indicated (Koriat, 1997). In other words, as participants learned more, they became less accurate in predicting their own learning.

However, the accuracy of metacognitive judgments may be influenced by the type of cue available at encoding. Specifically, some cues are known to influence both JOLs and memory performance while other cues are known to influence only JOLs but have no impact on actual memory, leading to metacognitive illusions. For example, among studies investigating the relationship between perceptual processing fluency, memory, and metacognition, it has been shown that while participants tend to assess stimuli with higher processing fluency (larger font, louder volume) as more likely to be recalled later, such perception has little to no effect on the participants' actual memory performance (Kuhlmann et al., 2020; Price et al., 2016; Rhodes & Castel, 2008, 2009). These illusions could be attributed to differences in perceived processing fluency or to participants' beliefs or prior experiences. Although it may seem that people expect to better remember large words solely due to their fluency (i.e., ease of processing), some research has found that words in larger fonts may be considered more important to remember than smaller fonts (Kornell et al., 2011; Li et al., 2015; Luna et al., 2018; Rhodes & Castel, 2008; see Chang & Brainerd, 2022 for a review). Other work on perceptual fluency and list composition revealed that memory effects were moderated by the design of the experiment (mixed list vs. pure list) and concluded that the font size effect is more common in mixed list designs, where small and large font size words are present simultaneously than in pure lists, where each list consisted of either only small words or only large words (Susser et al., 2013). Hence, font size is generally considered to be a cue that is not always diagnostic of actual memory performance but has a strong influence on metamemory.

Although most of the early work on the font size effect on JOLs and memory has mainly studied font size in isolation, recent work has looked at the degree to which participants integrate multiple cues when engaging in metamemory processes (Fan et al., 2021). To illustrate, it has been shown that people integrate font size with item relatedness (Price & Harrison, 2017; Rhodes & Castel, 2008), item value as manipulated through pairing word items with arbitrary numbers indicating their relative importance (Murphy et al., 2022), with font style (Price et al., 2016), as well as with the number of study presentations (Kornell et al., 2011; Undorf et al., 2018) to inform their JOLs. Moreover, it has been shown that font size, number of study presentations, concreteness, and emotionality affect JOLs not only at the aggregate level, but also at the individual level (Undorf et al., 2018). Together, the studies reveal that the degree to which font size informs JOLs and memory is conditional and depends on the other cues present at encoding. Specifically, Murphy et al. (2022) found that the magnitude of the effect of value was significantly greater than that of the font size, as high-value words were better recalled than low-value words, regardless of their perceptual fluency, and participants' metacognitive judgments mapped onto their performance. The present work aims to replicate and extend the prior work by Undorf et al. (2018) on the effect of emotional valence and font size on metacognition and recall performance. Such manipulation could provide novel insights into the theoretical framework that underlies cue-weighting (see Bröder & Undorf, 2019), such that people may differentially weigh and utilise cues available at encoding when one cue is known to have a greater impact on both metacognitive predictions and actual memory.

The current study

Previous work has shown that people generally predict to remember and remember emotional information better than neutral information (for a meta-analysis, see Yin et al., 2023). However, font size is known to lead to metacognitive illusions, such that people tend to predict a higher recall of large words than small words but often fail to demonstrate differences based on font size in their memory performance (for a recent meta-analysis, see Chang & Brainerd, 2022). Prior work on the simultaneous utilisation of multiple cues available at encoding in judgments of learning (JOLs) found that people integrate multiple extrinsic and intrinsic cues in JOLs when exposed to a single study-test session (Undorf et al., 2018; Undorf & Bröder, 2020).

The current work aims to replicate and further extend how people metacognitively incorporate emotional valence, which is known to elicit a strong effect on memory, and font size, which is known to elicit little to no effect on true memory. In other words, we examined whether the font size effect is robust to experience-based learning (through multiple study-test trials) when in competition with emotional valence that may influence JOLs and memory. Thus, in the current study participants were presented with lists of words varying in font size (small and large) and emotional valence (negative and neutral) and were asked to remember as many words as possible for a later recall test while engaging in item-level metacognitive assessments (JOLs in Experiment 1 and JOLs + restudy judgments in Experiment 2).

We aimed to investigate (1) whether people rely on both emotional valence and font size when making JOLs, (2) the effect of both cues on JOLs and recall performance during multi-trial learning, (3) whether emotional valence and/or font size guide control processes, such as restudy choices when multiple cues are available at encoding, and (4) whether participants' metacognitive assessments map onto their actual memory performance. Based on Koriat's (1997) cue-utilisation theory and relevant work by Undorf et al. (2018), we made the following predictions about the possible outcomes. We expected emotional valence to have an effect on both forms of metacognition (both JOLs and restudy judgments) as well as recall performance. We also expected font size to lead to metacognitive illusions, a dissociation between JOLs and actual recall, and the effect to be robust to multiple study-test trials.

Experiment 1

In Experiment 1, we investigated the effect of the font size and emotional valence of the to-be-remembered information on metacognition and recall by presenting participants with lists of words of varying in font size and emotional valence. Participants made JOLs after the presentation of each word and completed a free recall test after studying each list. We expected participants to incorporate both font size and emotional valence when making JOLs as both cues are known to influence metacognitive monitoring processes when presented in isolation. Specifically, we expected higher JOLs for emotional words than neutral words, as well as higher JOLs for large words than small words. However, based on prior work, we predicted font size to affect JOLs but not actual memory while emotional valence to affect both JOLs and actual recall performance. Overall, we hypothesised a strong effect of emotional valence on recall and metacognition. We also hypothesised that while in the presence of both font size and emotional valence, participants should rely more on emotional valence, which is known to be diagnostic of subsequent memory performance, JOLs may be influenced by both cues.

Method

Participants. Participants were 47 undergraduate students (age: $M = 20.44$, $SD = 1.30$) recruited from the University of California, Los Angeles (UCLA) Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from the analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire, which resulted in one exclusion (participants received credit for their participation regardless of their answer). An a priori power analysis indicated that for a 2 (emotional valence: neutral, negative) × 2 (font size: small, large) repeated-measures, within-subjects ANOVA, assuming alpha = .05 and power = .80, 50 participants would be needed to reliably detect a small to medium effect size ($\eta^2 = .04$).

Materials. The stimuli were 96 nouns of varying font sizes (i.e., small and large) and emotional valences (i.e., negative and neutral) adapted from Eich and Castel (2016) (see Appendix A). Since previous findings showing a memory advantage for emotional over non-emotional words demonstrate a stronger effect for negatively valenced items due to their increased distinctiveness at encoding (Ochsner, 2000), we included only negative words. Half of the words were negatively valenced ($M_{valence} = 2.8$, $M_{arousal} = 4.4$, e.g., killer, murder) while the other half were neutrally valenced ($M_{valence} = 6.0$, $M_{arousal} = 4.6$, e.g., train, basket). Within the negatively valenced words, half (i.e., 24) were of small font size (12-point Arial font) while the other half were of large font size (48-point Arial font). Similarly, within the neutrally valenced words, half (i.e., 24) were of small font size (12-point Arial font) while the other half were of large font size (48-point Arial font). Words had a length of between three and eight letters ($M = 5.0$, $SD = 1.08$) and their concreteness level (on a scale from 1 to 5, with 5 representing the highest concreteness level; $M = 4.43$, $SD = .51$) did not vary between small ($M = 4.47$, $SD = .50$) and large ($M = 4.39$, $SD = .53$) font size words [$t(94) = -.86$, $p = .395$]. However, concreteness level varied between negative ($M = 4.20$, $SD = .52$) and neutral ($M = 4.66$, $SD = .39$) words [Welch corrected results: $t(89.15) = -4.79$, $p < .001$]. The words were classified according to the English Lexicon Project website (Balota et al., 2007).

Lists were constructed using a mixed-list design, such that a single list consisted of emotional and neutral words, as well as small and large words. A total of four lists, each containing 24 words with an equal number of items from each condition were created (see Appendix A). Specifically, each list contained six unique negative-small font words, six negative-large font words, six neutral-small font words, and six neutral-large words. Each participant studied all four lists of 24 words during the experiment, and word order was entirely randomised for each list. The stimuli have been made available on the Open Science Framework: https://osf.io/9g4nb/.

Procedure. Participants were told that they would study lists of words and be asked to recall as many words as they could remember after studying each list. The words were then displayed on the participant's computer screen, one at a time, for three seconds each. After each word was presented, participants were instructed to make JOLs to estimate the probability of recalling it on a later test and were given five seconds to record their judgment. Participants responded with a number between 0 and 100, with 0 meaning they believed they would most likely not remember the word and 100 meaning they believed they would most likely remember the word. After the presentation of all 24 words within each list, participants were given a two-minute free-recall test to recall as many words as they could from the list they just studied. Immediately following the recall period, participants were told how many words they recalled from a given list and repeated the process for a total of four study-test blocks.

Results

Judgments of learning (JOLs). Figure 1 presents the judgments of learning (JOLs) as a function of font size and emotional valence. To investigate any differences in JOLs based on font size, emotional valence, and task experience, a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) × 4 (list: 1, 2, 3, 4) repeated measures ANOVA was performed and revealed a main effect of emotional valence [$F(1, 46) = 23.81$, $p < .001$, $\eta^2 = .12$] such that the negative words ($M = 52.05$, $SD = 21.56$) were given higher JOLs than the neutral words ($M = 43.62$, $SD = 19.90$). Additionally, the results revealed a main effect of font size [$F(1, 46) = 14.19$, $p < .001$, $\eta^2 = .02$] such that the large words ($M = 49.45$, $SD = 20.80$) were given higher JOLs than the small words ($M = 46.21$, $SD = 20.63$). Next, there was a main effect of list [Mauchly's $W = .59$, $p < .001$, Huynh-Feldt corrected results: $F(2.33, 107.26) = 3.08$, $p = .043$, $\eta^2 = .02$], which revealed that JOLs slightly decreased with task experience. The results also revealed a significant interaction between emotional valence and list [Mauchly's $W = .73$, $p = .014$, Huynh-Feldt corrected results: $F(2.60, 119.65) = 13.24$, $p < .001$, $\eta^2 = .02$], such that JOLs for negative words decreased with task experience and increased for neutral words. However, font size did not interact with emotional valence [$F(1, 46) = 1.18$, $p = .284$, $\eta^2 < .01$] and did not interact with list [$F(3, 138) = .40$, $p = .752$, $\eta^2 < .01$]. The three-way interaction (font size × emotional valence × list) was also not significant [$F(3, 138) = .16$, $p = .922$, $\eta^2 < .01$].

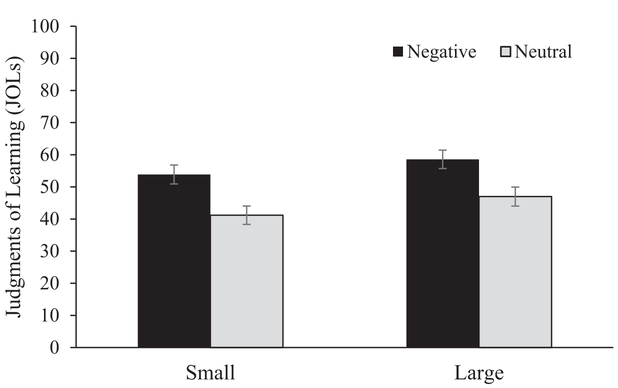

Figure 1. Judgments of learning (JOLs) as a function of font size and emotional valence in Experiment 1. Error bars reflect the standard error of the mean.

Furthermore, since prior research on concreteness revealed an influence on individuals' beliefs about word memorability (Witherby & Tauber, 2017), we utilised multilevel models (MLMs), also known as mixed-effects or hierarchical models, to demonstrate the degree to which concreteness is related to the other factors. We approached the data as hierarchical or clustered, employing a multilevel framework where items were nested within individual participants. In the context of linear models, we adopted restricted maximum likelihood estimation (REML) to determine coefficients. This method was chosen for its robustness in handling small sample sizes at level-2, corresponding to the participant level in our study (McNeish, 2017).

The model revealed significant main effects for emotional valence (Est. = −8.32, SE = .63, $t = -13.27$, $p < .001$), font size (Est. = 3.04, SE = .63, $t = 4.85$, $p < .001$) and no significant main effect for concreteness (Est. = 3.04, SE = .63, $t = 4.85$, $p < .001$). Hence, the results confirm ANOVA findings. However, a significant interaction was observed between valence and concreteness (Est. = 3.11, SE = 1.24, $t = 2.52$, $p = .012$), indicating that the relationship between emotional valence and JOLs varied depending on concreteness, such that for negative words only, more concrete words received lower JOLs, demonstrating a more nuanced relationship between valence and metacognitive judgments. No other significant interactions were observed.

Recall performance. Figure 2 shows the results on recall as a function of font size and emotional valence. To investigate any differences in recall performance based on font size, emotional valence, and task experience, a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) × 4 (list: 1, 2, 3, 4) repeated measures ANOVA was performed and revealed a main effect of emotional valence [$F(1, 46) = 16.34$, $p < .001$, $\eta^2 = .05$] such that the negative words ($M = .60$, $SD = .21$) were recalled better than the neutral words ($M = .51$, $SD = .23$). However, results revealed no main effect of font size [F(1,

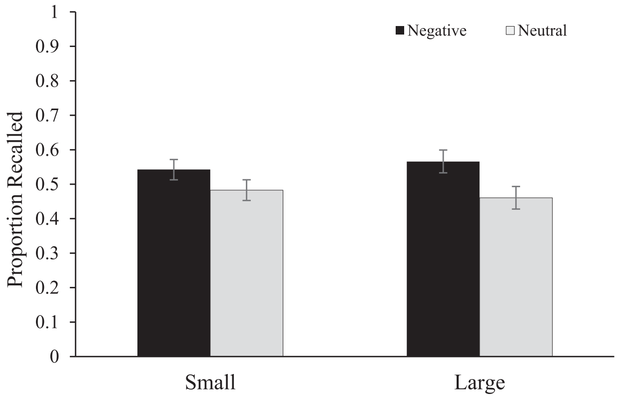

Figure 2. Recall performance as a function of font size and emotional valence in Experiment 1. Error bars reflect the standard error of the mean.

46) = 4.08, $p = .050$, $\eta^2 < .01$], such that participants' performance did not significantly differ between large ($M = .57$, $SD = .22$) and small ($M = .54$, $SD = .21$) words. Next, there was not a main effect of list [$F(3, 138) = 2.63$, $p = .053$, $\eta^2 < .01$], which revealed that recall performance did not change with task experience. Although we found no significant interaction between emotional valence and font size [$F(1, 46) = .41$, $p = .528$, $\eta^2 < .01$], as well as between font size and list [$F(3, 138) = 1.11$, $p = .348$, $\eta^2 < .01$], the results revealed a significant interaction between emotional valence and list [$F(3, 138) = 3.57$, $p = .016$, $\eta^2 = .01$], such that the recall for neutral words decreased with task experience and increased for negative words. Lastly, the three-way interaction (font size × emotional valence × list) was also not significant [$F(3, 138) = .919$, $p = .434$, $\eta^2 < .01$].

We then performed a logistic MLM with recall accuracy (level 1) modelled as a function of font size, emotional valence, and concreteness. Because memory performance at the item-level was binary (i.e., correct or incorrect), we conducted a logistic MLM to assess performance. The results revealed that valence significantly predicted accuracy [exp(B) = .66, CI: .58 – .74, $z = -6.51$, $p < .001$], such that negative words were recalled better than neutral words. No other main effects or interactions were significant.

Metacognitive accuracy. We then calculated measures of *relative accuracy*[1] (i.e., *resolution*) to detect whether participants were able to discriminate between words that would be remembered and those that would not (Rhodes, 2016). A strong positive correlation between JOLs and recall performance would indicate that a participant remembers what they say they will remember while a strong negative correlation would indicate that a participant forgets what they say they will remember. Gamma correlations for each participant were calculated to determine relative resolution as a function of font size and emotional valence. A repeated measures ANOVA (4 levels: list 1, list 2, list 3, list 4) did not reveal a main effect of list [Mauchly's $W = .57$, $p < .001$, Huynh-Feldt corrected results: $F(2.36, 89.62) = 1.25$, $p = .295$, $\eta^2 = .03$], indicating that relative accuracy did not change with task experience. We also examined Gamma correlations for each participant as a function of emotional valence and font size using a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) repeated-measures ANOVA. Results revealed that participants' relative accuracy did not significantly differ between large font size words ($M = .29$, $SD = .40$) and small font size words ($M = .32$, $SD = .35$), [$F(1, 41) = .16$, $p = .691$, $\eta^2 < .01$]. Additionally, participants' relative accuracy did not significantly differ between negative words ($M = .29$, $SD = .38$) and neutral words ($M = .32$, $SD = .35$), [$F(1, 41) = .24$, $p = .629$, $\eta^2 < .01$]. Font size did not interact with emotional valence [$F(1, 41) = .49$, $p = .486$, $\eta^2 < .01$].

Discussion

In Experiment 1, we examined the effect of font size and emotional valence on JOLs and recall performance. Consistent with prior literature, participants expected perceptual features of visually presented stimuli (i.e., font size) to influence their memory by giving higher JOLs to words of large font size than words of small font size, but their actual memory performance did not reveal any significant differences between the varying font sizes (Rhodes & Castel, 2008). Although font size is generally not indicative of actual memory, participants tend to incorporate font size when making predictions of future memory as measured by item-level JOLs, even when emotional valence, a cue well known for its effect on recall, is also present. Furthermore, we observed that the relationship between valence and metacognitive judgments is nuanced. Although we found no main effect of concreteness on JOLs, the interaction between valence and concreteness was significant, such that for negative words only, more concrete words received lower JOLs. Future research should include a broader spectrum of emotionally negative words, varying in concreteness, to validate and elucidate these unexpected results.

Given that emotional valence has an effect on both JOLs and recall performance, it is indicative of actual memory performance (Hourihan et al., 2017; Kensinger & Corkin, 2003; Mathews & Barch, 2004), and continues to demonstrate a strong effect when competing with cues that are known to elicit little to no effect on actual memory. In other words, because emotional information tends to engage more semantic or conceptual processing during the encoding phase, it is likely that, unlike more perceptual cues (i.e., font size), emotional valence engages deeper levels of processing and greater distinctiveness leading to better memory (Jay et al., 2008). Lastly, in terms of relative accuracy, participants were metacognitively accurate in their JOLs and generally recalled words they judged they would remember.

Based on prior literature on the utilisation of multiple cues in JOLs, when both font size and number of study presentations (extrinsic cue) were manipulated at encoding, both cues affected JOLs. However, while both cues influenced memory performance, the effect of emotional valence on recall was much larger than that of font size (Undorf et al., 2018). Additionally, research has shown that when multiple cues, such as word concreteness, number of study presentations, emotionality, and font size are available at encoding, all cues were shown to influence JOLs, such that JOLs increase with increasing font size, concreteness, and emotionality. However, unlike concreteness and emotionality, font size had only a small effect on recall performance, which further suggests that font size has little to no effect on memory (Undorf et al., 2018). Finally, research has shown that when concreteness and emotionality were experimentally manipulated at encoding, both cues tended to influence JOLs and actual recall performance, which further proves that emotional valence has an effect on memory even when presented along with other cues (Undorf et al., 2018). Hence, Experiment 1 adds to the existing knowledge on cue utilisation and suggests that at the aggregate level, when both emotional valence and font size are available at encoding, people rely on both cues when making JOLs but only emotional valence predicts recall performance. Moreover, Experiment 1 extended prior literature on multi-trial learning and revealed that the font size effect on JOLs and memory is robust, such that participants continue to demonstrate a dissociation between predicted and actual memory throughout multiple study-test cycles with new materials.

Experiment 2

In Experiment 1, the results revealed that while both font size and emotional valence influence metacognition, only emotional valence is indicative of actual memory performance. Thus, when both emotional valence and font size are present at encoding, participants integrate the two cues when making JOLs despite their differential effect on memory. Previous literature has also shown that metacognitive monitoring plays a critical role in control processes, such as strategic regulation of information processing and behaviour (Nelson, 1996). For example, according to the discrepancy reduction theory, when learning new information, learners seek to reduce the discrepancy between their current and desired or optimal state of learning (Dunlosky & Hertzog, 1998; Rhodes & Castel, 2009). Specifically, it has been shown that learners tend to allocate more of their total study time to items they think as being more difficult to encode when compared to easier items.

In Experiment 2, we aimed to extend the previous findings (Experiment 1) and investigate the effect of font size and emotional valence on metacognitive control processes involved in the regulation of study behaviour. Specifically, after each word was presented, we asked participants to make a JOL and decide whether they would want another opportunity to restudy each word they were presented with. We expected participants to more frequently choose to restudy the neutral words compared to negative words, as well as small font size words compared to large font size words because neutral and small words are typically perceived as being less salient or less easily processed than negative and large words. We also expected a negative correlation between JOLs and restudy judgments, such that words that received lower JOLs would be chosen for restudy more often than words that received higher JOLs. Hence, this finding would further extend the discrepancy reduction model to instances when multiple cues are available at encoding, where one of the cues is known to be a better indicator of actual memory performance.

Method

Participants. Participants were 48 undergraduate students (age: $M = 20.5$, $SD = 2.96$) recruited from the UCLA Human Subjects Pool who received course credit for their participation. The experiment was conducted online and lasted approximately 30 min. Participants were excluded from the analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire. This exclusion process resulted in one exclusion.

Materials and procedure. The materials in Experiment 2 were identical to those in Experiment 1. The procedure was very similar as in Experiment 1, in that words were presented for three seconds each, and participants were instructed to recall as many words as they could from each of four lists of 24 words each. However, after studying each word, participants made both JOLs and restudy judgments by indicating whether they would like the opportunity to restudy each item later. Participants were presented with the restudy judgment along with a YES and NO option and had to choose one to proceed to the next word. However, they were not given the opportunity to restudy items as this could influence recall (see Rhodes & Castel, 2009, for a similar procedure). Immediately following the recall period, participants were told how many words they recalled from a given list and repeated the process for a total of four study-test blocks.

Results

Judgments of learning (JOLs). Figure 3 shows the judgments of learning (JOLs) as a function of font size and emotional valence. To investigate possible differences in JOLs based on font size, emotional valence, and task experience, a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) × 4 (list: 1, 2, 3, 4) repeated measures ANOVA was performed and revealed a main effect of emotional valence [$F(1, 42) = 49.33$, $p < .001$, $\eta^2 = .15$] such that the negative words ($M = 54.21$, $SD = 19.62$) were given higher JOLs than the neutral words ($M = 44.04$, $SD = 20.75$). Additionally, the results revealed a main effect of font size [$F(1, 42) = 10.48$, $p = .002$, $\eta^2 = .03$] such that the larger words ($M = 50.45$, $SD = 20.04$) were given higher JOLs than the smaller words

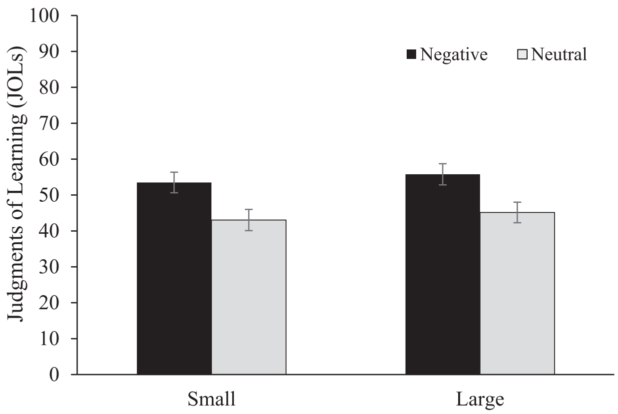

Figure 3. Judgments of learning (JOLs) as a function of font size and emotional valence in Experiment 2. Error bars reflect the standard error of the mean.

($M = 47.80$, $SD = 20.32$). Next, there was a main effect of list [Mauchly's $W = .23$, $p < .001$, Huynh-Feldt corrected results: $F(1.72, 72.02) = 4.58$, $p = .018$, $\eta^2 = .04$], which revealed that JOLs slightly decreased with task experience. The results also revealed a significant interaction between emotional valence and list [$F(3, 126) = 13.31$, $p < .001$, $\eta^2 = .03$], such that JOLs for negative words decreased with task experience and increased for neutral words. Additionally, font size interacted with list [$F(3, 126) = 3.60$, $p = .015$, $\eta^2 = .01$] and emotional valence [$F(1, 42) = 4.20$, $p = .047$, $\eta^2 < .01$], such that with task experience, JOLs for smaller and negative words decreased significantly more than for larger and neutral words. The three-way interaction (font size × emotional valence × list) was not significant [$F(3, 126) = 1.91$, $p = .132$, $\eta^2 < .01$].

Furthermore, we computed the correlations between concreteness and JOLs ($r = -.15$, $p < .001$) as well as concreteness and recall ($r = --.06$, $p < .001$) from Experiment 2. The results of the regression analysis investigating the joint impact of concreteness, emotional valence, and font size on JOLs revealed that the model explained a significant amount of variance in JOLs [$R^2 = .04$, adjusted $R^2 = .04$, $F(4, 4054) = 41.13$, $p < .001$]. Specifically, font size ($\beta = -.08$, $t = -2.55$, $p = .014$), emotional valence ($\beta = -.29$, $t = -8.23$, $p < .001$), and concreteness ($\beta = -.08$, $t = -4.55$, $p < .001$) were negatively associated with JOLs, such that larger, more negative, and less concrete words received higher JOLs.

We then utilised MLM to demonstrate the degree to which concreteness is related to emotional valence and font size. The model revealed significant main effects for emotional valence (Est. = −8.32, SE = .78, $t = -10.70$, $p < .001$), font size (Est. = −2.09, SE = .78, $t = -13.27$, $p < .001$), and concreteness (Est. = −4.28, SE = .63, $t = 4.85$, $p < .001$). Hence, the results confirm ANOVA findings. However, a significant interaction was observed between valence and concreteness (Est. = 5.87, SE = 1.57, $t = 3.79$, $p < .001$), indicating that the relationship between emotional valence and JOLs varied depending on concreteness, such that for negative words only, more concrete words received lower JOLs, as was found in Experiment 1, but this may be due to a restricted range of words used in the present study, and future research could examine this issue in more detail with a more suitable design. No other significant interactions were observed.

Restudy judgments. Restudy choices as a function of font size and emotional valence are shown in Figure 4. To investigate possible differences in restudy choices based on font size, emotional valence, and task experience, a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) × 4 (list: 1, 2, 3, 4) repeated measures ANOVA was performed and revealed a main effect of emotional valence [$F(1, 41) = 21.23$, $p < .001$, $\eta^2 = .06$] such that participants chose to restudy a reliably larger percentage of neutral words ($M = .37$, $SD = .33$) in comparison with negative words ($M = .27$, $SD = .29$). Further, the results revealed a small main effect of font size [$F(1, 41) = 4.88$, $p = .033$, $\eta^2 < .01$] such that participants choose to restudy the smaller words ($M = .36$, $SD = .32$) more than the larger words ($M = .29$, $SD = .30$). Next, there was not a main effect of list [Mauchly's $W = .22$, $p < .001$, Huynh-Feldt corrected results: $F(1.60, 65.51) = 1.76$, $p = .185$, $\eta^2 = .02$], which revealed that the rate of restudy judgments did not change with task experience. Although the results revealed that font size did not interact with emotional valence [$F(1, 41) = .09$, $p = .772$, $\eta^2 < .01$] as well as with list [$F(3, 123) = .12$, $p = .950$, $\eta^2 < .01$], emotional valence interacted with list [$F(3, 123) = 14.55$, $p < .001$, $\eta^2 = .03$], such that with task experience, participants chose to restudy more of the neutral words than negative words. The three-way interaction (font size × emotional valence × list) was also not significant [$F(3, 123) = 1.48$, $p = .223$, $\eta^2 < .01$]. Together, the results revealed that JOLs were sensitive to both font size and emotional valence of the studied items while restudy choices were mostly sensitive to emotional valence.

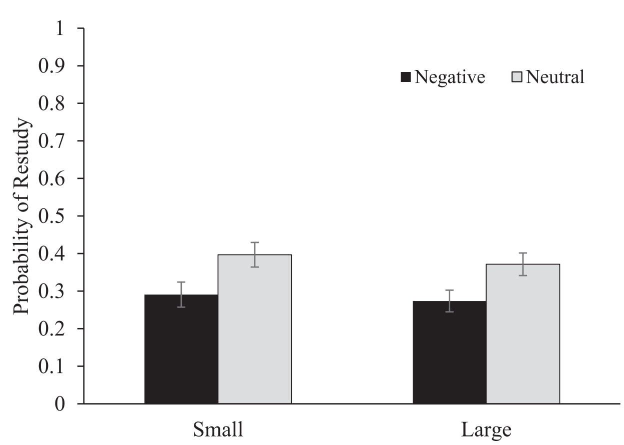

Figure 4. Restudy judgments as a function of font size and emotional valence in Experiment 2. Error bars reflect the standard error of the mean.

We then performed a logistic MLM with restudy choices (level 1) modelled as a function of font size, emotional valence, and concreteness. Because restudy judgments at the item-level were binary (i.e., correct or incorrect), we conducted a logistic MLM to assess performance. The results revealed that valence significantly predicted restudy judgments [eB = 2.10, CI: 1.73–2.55, $z = 7.47$, $p < .001$], such that neutral words were chosen for restudy more than negative words. Additionally, there was a main effect of font size [exp(B) = 1.22, CI: 1.00–1.48, $z = 2.00$, $p = .046$], such that small words were chosen for restudy more than large words. No other main effects or interactions were significant.

Recall performance. Figure 5 displays the results of recall as a function of font size and emotional valence. To investigate possible differences in recall performance based on font size, emotional valence, and task experience, a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) × 4 (list: 1, 2, 3, 4) repeated measures ANOVA was performed and revealed a main effect of emotional valence [$F(1, 47) = 24.74$, $p < .001$, $\eta^2 = .05$] such that the negative words ($M = .53$, $SD = .19$) were recalled better than the neutral words ($M = .44$, $SD = .19$). Additionally, results revealed a small main effect of font size [$F(1, 47) = 6.98$, $p = .011$, $\eta^2 = .01$], such that the large font size words ($M = .50$, $SD = .19$) were recalled better than the small font size words ($M = .46$, $SD = .19$). Next, there was a main effect of list [Mauchly's $W = .72$, $p = .010$, Huynh-Feldt corrected results: $F(2.64, 124.22) = 3.22$, $p = .030$, $\eta^2 = .02$], which revealed that recall performance decreased with task experience. The results also revealed a significant interaction between emotional valence and list [$F(3, 141) = 3.52$, $p = .017$, $\eta^2 = .01$], such that the recall for negative words decreased with task experience and increased for neutral words. However, font size did not interact with emotional valence [$F(1, 47) = .79$, $p = .382$, $\eta^2 < .01$] and did not interact with list [$F(3, 141) = 1.67$, $p = .177$, $\eta^2 = .01$]. The three-way interaction (font size × emotional valence × list) was also not significant [$F(3, 141) = 0.43$, $p = .729$, $\eta^2 < .01$].

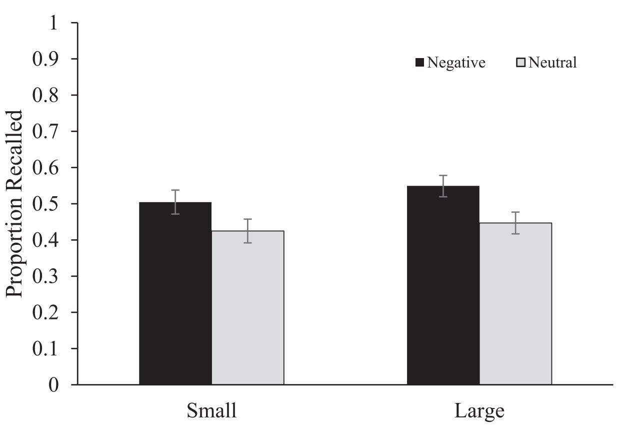

Figure 5. Recall performance as a function of font size and emotional valence in Experiment 2. Error bars reflect the standard error of the mean.

We then performed a logistic MLM with recall accuracy (level 1) modelled as a function of font size, emotional valence, and concreteness. Because memory performance at the item-level was binary (i.e., correct or incorrect), we conducted a logistic MLM to assess performance. The results revealed that valence significantly predicted accuracy [exp(B) = .02, CI: .01 – .03, $z = -17.35$, $p < .001$], such that negative words were recalled better than neutral words. Moreover, font size [eB = .15, CI: .10–.23, $z = -8.33$, $p < .001$] as well as concreteness [eB = .39, CI: .25–.61, $z = -4.16$, $p < .001$] significantly predicted accuracy, such that large font size and more concrete words were recalled better than small font size and less concrete words.

There were also several significant interactions: valence and font size [eB = 25.87, CI: 10.54–63.54, $z = 7.10$, $p < .001$], valence and concreteness [eB = 4.77, CI: 1.95–11.65, $z = 3.43$, $p < .001$], as well as font size and concreteness [eB = .03, CI: .01–.07, $z = -7.81$, $p < .001$]. The results indicate that the relationship between valence and recall accuracy varied depending on font size, such that for negative words only, larger words were recalled better. Additionally, the relationship between valence and recall accuracy varied depending on concreteness, such that for negative words only, more concrete words were recalled better. Finally, the relationship between font size and recall accuracy varied depending on concreteness, such that for large words only, more concrete words were recalled better.

Metacognitive accuracy. A repeated measures ANOVA (4 levels) revealed no main effect of list [$F(3, 114) = 2.68$, $p = .050$, $\eta^2 = .07$], indicating that relative accuracy did not change with task experience. We also examined Gamma correlations for each participant as a function of emotional valence and font size using a 2 (emotional valence: negative, neutral) × 2 (font size: large, small) repeated-measures ANOVA. Results revealed that participants' relative accuracy did not significantly differ between large font size words ($M = .26$, $SD = .38$) and small font size words ($M = .24$, $SD = .41$), [$F(1, 41) = .05$, $p = .828$, $\eta^2 < .01$]. Additionally, participants were similarly relatively accurate for negative words ($M = .30$, $SD = .38$) and neutral words ($M = .20$, $SD = .41$), [$F(1, 41) = 2.01$, $p = .164$, $\eta^2 = .02$]. Font size did not interact with emotional valence [$F(1, 41) = .09$, $p = .769$, $\eta^2 < .01$]. Additionally, gamma correlations between a given word's JOL and whether the word was chosen to be restudied were calculated for each participant, and these correlations ($M = -.46$, $SD = .45$) were different than 0 [one sample t-test: $t(40) = -6.50$, $p < .001$, $d = -1.02$]. Hence, participants were relatively accurate in their JOLs and restudy judgments and tended to choose to restudy items they expected to forget.

Discussion

The results of Experiment 2 revealed that while participants based their metacognitive monitoring (JOLs) on both font size and emotional valence, their control processes (restudy choices) were generally influenced by emotional valence. Participants may more frequently choose to restudy words they expected to forget or found difficult to remember, and emotional valence showed a stronger effect on self-regulation of study behaviour. Participants predicted a higher recall of negative words than neutral words and later wanted to restudy the neutral words, likely because they based their control processes on metacognitive monitoring. It may be that if one expects to better remember negative words and likely forget neutral words, an effective strategy would be to restudy the neutral words to ensure optimal learning, and thus, reduce the discrepancy between the current and optimal state of learning (Dunlosky & Hertzog, 1998). However, although participants expected to remember small font size words less well than large font size words as indicated through their JOLs, they still chose to restudy a slightly higher proportion of small words over large words. Overall, participants utilised both font size and emotional valence when making JOLs such that larger words and negative words were given higher JOLs than small words and neutral words, respectively. However, unlike in Experiment 1, participants' recall was sensitive to both font size and emotional valence, as some prior work shows that font size can influence recall to a small degree (Chang & Brainerd, 2022). This could be due to beliefs about the perceived importance of emotional words and larger font size (cf., Blake & Castel, 2018; Luna et al., 2019; Yang et al., 2018), and the metacognitive judgments that were made in Experiment 2, although additional research would be needed to examine this issue in more detail.

General discussion

We often encounter perceptual information that varies in emotional valence and font size, such as when we read a news article or see product names in advertisements. Generally, people perceive emotional information as more memorable than neutral (i.e., non-emotional) information and tend to better remember emotional information (for a meta-analysis, see Yin et al., 2023). While certain stimuli properties are known to influence both predicted and actual memory when presented in isolation, less is known about the combined effect of multiple cues available at encoding when one cue is more valid for judging memorability. To illustrate, news headlines use multiple visual cues, such as emotionality and font size to influence readers' attention and memory. When information is neutral, font size in isolation often has little to no influence on memory performance despite people's metacognitive beliefs about large font size affecting memory (Blake & Castel, 2018; Chang & Brainerd, 2022; Luna et al., 2018; Mueller et al., 2014; Rhodes & Castel, 2008).

In the current study, participants were presented with lists of words varying in font size (small and large) and emotional valence (negative and neutral). After studying each word, participants either provided JOLs only (Experiment 1) or made both JOLs and restudy judgments by choosing whether they would want to restudy a given word (Experiment 2). Generally, the results further support the idea that even in the presence of a more valid cue, people tend to expect a higher recall of large font size words compared to small font size words, even when their actual memory fails to show any differences based on font size (Rhodes & Castel, 2008; 2009), and the current study revealed that the font size effect on JOLs and memory is robust to multi-trial learning. Thus, the present study is consistent with the idea of cue-weighting (Bröder & Undorf, 2019; Koriat, 1997; Undorf et al., 2018; Undorf & Bröder, 2020), suggesting that multiple cues inform metacognitive judgments. Specifically, participants tend to rely on both font size and emotional valence when making JOLs and are accurate in their JOLs, but only emotional valence shows an effect on memory. Given that emotional valence is known to be diagnostic of later recall, its effect remains strong even in the presence of a less valid cue known to lead to metacognitive illusions, which is consistent with the emotion saliency effect (Fung & Carstensen, 2003) and emotion saliency effect on JOLs (Tauber et al., 2017; Witherby et al., 2021; Zimmerman & Kelley, 2010).

Additionally, the current work provides novel insights into how font size and emotional valence influence control processes, such as self-regulation of study behaviour. Experiment 2 revealed that consistent with Experiment 1, participants relied on both cues when engaging in JOLs; however, their restudy choices were generally influenced by emotional valence. It may be that participants expected to forget neutral words and small words, which is consistent with the discrepancy reduction model (Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 2004), but generally chose to restudy the negative words, showing no strong restudy preference based on font size when these two cues are in competition. It may be that study decisions engage processes that allow one to focus more on the cues that actually impact memory and less on cues that do not impact memory, such that in the present study, control processes may be more tuned into what is actually influencing memory. Thus, control processes can selectively prioritise emotionally salient stimuli for further study, while disregarding less relevant cues such as font size, which provides valuable insights into how individuals strategically allocate their cognitive resources during the learning process (Koriat, 1997).

Together, the results show that in instances when both font size and emotional valence are present at encoding, emotional stimuli have an effect on both metacognitive monitoring and control processes while font size only influences metacognitive monitoring and has little influence on recall. Hence, by considering multiple cues during the learning phase, individuals may be better equipped to differentiate between cues that genuinely impact memory (emotional valence) and cues that do not (font size). Therefore, when multiple cues, such as both emotional valence and font size are available at encoding, engaging in both monitoring and control assessments can potentially reduce metacognitive illusions. This finding contributes to our understanding of how individuals optimise their metacognitive accuracy when competing cues (emotional valence, font size) are present during encoding.

However, there were a few methodological limitations in the present work. One of the limitations of Experiment 2 was that participants were asked to make JOLs and restudy decisions within a single trial (after studying each of the words). While we acknowledge that this setup may raise concerns about potential consistency and demand effects, prior work has widely used a similar procedure (Nelson et al., 1994; Rhodes, 2019; Rhodes & Castel, 2009). By utilising this established approach, we aimed to build upon the existing body of literature and provide continuity for the current investigation. Although we think that JOLs may inform restudy choices, it is unclear how the manipulated variables and/or the JOLs may influence this directly (see Metcalfe & Finn, 2008 for an example of how accurate monitoring can lead to study choices that enhance learning). Future studies could delve further into this potential limitation by adopting alternative methodologies. For instance, researchers might consider separating the JOL and restudy decision phases, allowing participants more time for reflection between these two tasks.

Another potential limitation of Experiment 2 was the absence of an opportunity for participants to restudy the words they had selected for restudy. It is important to note that our primary focus in this study was centered around investigating the effect of font size and emotional valence on metacognitive judgments and control processes, rather than memory effects. As such, we deliberately chose to avoid the potential confounding effect that honoring restudy choices might introduce on free recall performance. Nevertheless, our results demonstrate that the rate of restudy decisions did not decline across lists suggesting that participants' restudy decisions were likely not impacted by the awareness that these decisions resulted in restudy opportunities. However, future research endeavours could take a different approach by honouring participants' restudy choices and then assessing both memory performance and metacognitive accuracy. It may also be interesting and informative to examine these effects using more complex emotional materials (such as images) and if other populations (such as older adults) may show similar effects, suggesting that emotional information can strongly guide metacognition even in the presence of other competing environmental cues.

Note

1. The experiments reported in this article were formally preregistered. The stimuli and data have been made available on the Open Science Framework https://osf.io/9g4nb/.The inclusion of relative accuracy, though not initially mentioned in the original preregistered plans, was later updated. This inclusion did not significantly impact our original conclusions, which remain consistent with the study's primary objectives.

Open Scholarship

This article has earned the Center for Open Science badges for Open Data, Open Materials and Preregistered. The data and materials are openly accessible at https://osf.io/9g4nb/.

Acknowledgments

We would like to thank Feifan Zou, Nicole Elbaz-Deckel, and Zoe Liau for their help in data collection and data coding. We also thank Dillon H. Murphy, Mary C. Whatley, and Ashley Chen for providing useful feedback regarding the manuscript as well as for their help in the review process.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The experiments reported in this study were formally preregistered. The stimuli and data have been made available on the Open Science Framework at https://doi.org/10.17605/OSF.IO/9G4NB.

ORCID

Karina Agadzhanyan http://orcid.org/0000-0002-1938-6068
Alan D. Castel http://orcid.org/0000-0003-1965-8227

References

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

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief. *Acta Psychologica*, *186*, 27–38. https://doi.org/10.1016/j.actpsy.2018.04.004

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica*, *197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Chang, M., & Brainerd, C. J. (2022). Association and dissociation between judgments of learning and memory: A meta-analysis of the font size effect. *Metacognition and Learning*, 1–34.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Erlbaum.

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (pp. 23–37). Oxford University Press.

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory & Cognition*, *32*(5), 779–788. https://doi.org/10.3758/BF03195868

Eich, T. S., & Castel, A. D. (2016). The cognitive control of emotional versus value-based information in younger and older adults. *Psychology & Aging*, *31*(5), 503. https://doi.org/10.1037/pag0000106

Fan, T., Zheng, J., Hu, X., Su, N., Yin, Y., Yang, C., & Luo, L. (2021). The contribution of metamemory beliefs to the font size effect on judgments of learning: Is word frequency a moderating factor? *PLoS One*, *16*, e0257547.

Fung, H. H., & Carstensen, L. L. (2003). Sending memorable messages to the old: Age differences in preferences and memory for advertisements. *Journal of Personality and Social Psychology*, *85*(1), 163–178. https://doi.org/10.1037/0022-3514.85.1.163

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition*, *45*(1), 121–136. https://doi.org/10.3758/s13421-016-0646-3

Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *The American Journal of Psychology*, *121*(1), 83–103. https://doi.org/10.2307/20445445

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*(2), 99–113. https://doi.org/10.1177/1754073908100432

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

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Kuhlmann, B. G., Frank, D. J., & Danner, D. (2020). Louder = larger = clearer. *Zeitschrift für Psychologie*, *228*(4), 296–300. https://doi.org/10.1027/2151-2604/a000427

Li, F., Xie, R., Li, X., & Li, W. (2015). The influence of perceptual information on control processes involved in self-regulated learning: Evidence from item selection. *Psychonomic Bulletin & Review*, *22*(4), 1007–1013. https://doi.org/10.3758/s13423-014-0762-7

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology*, *71*(7), 1626–1636. https://doi.org/10.1080/17470218.2017.1343362

Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: Explaining the belief that font size affects memory. *Memory*, *27*(4), 555–560. https://doi.org/10.1080/09658211.2018.1529797

Mathews, J., & Barch, D. (2004). Episodic memory for emotional and nonemotional words in schizophrenia. *Cognition & Emotion*, *18*(6), 721–740. https://doi.org/10.1080/02699930341000284

McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward–Roger correction. *Multivariate Behavioral Research*, *52*, 661–670.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. https://doi.org/10.3758/PBR.15.1.174

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Murphy, D. H., & Castel, A. D. (2022). Selective remembering and directed forgetting are influenced by similar stimulus properties. *Memory*, *30*(9), 1130–1147. https://doi.org/10.1080/09658211.2022.2092152

Murphy, D. H., Huckins, S. C., Rhodes, M. G., & Castel, A. D. (2022). The effect of perceptual processing fluency and value on metacognition and remembering. *Psychonomic Bulletin & Review*, *29*(3), 910–921. https://doi.org/10.3758/s13423-021-02030-8

Murphy, N. A., & Isaacowitz, D. M. (2008). Preferences for emotional information in older and younger adults: A meta-analysis of memory and attention tasks. *Psychology and Aging*, *23*(2), 263. https://doi.org/10.1037/0882-7974.23.2.263

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*(2), 102. https://doi.org/10.1037/0003-066X.51.2.102

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*(4), 207–213. https://doi.org/10.1111/j.1467-9280.1994.tb00502.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 80, 125–173). Academic Press.

Ochsner, K. N. (2000). Are affective events richly "remembered" or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, *129*(2), 242–261. https://doi.org/10.1037/0096-3445.129.2.242

Price, J., & Harrison, A. (2017). Examining what prestudy and immediate judgments of learning reveal about the bases of metamemory judgments. *Journal of Memory and Language*, *94*, 177–194. https://doi.org/10.1016/j.jml.2016.12.003

Price, J., McElroy, K., & Martin, N. J. (2016). The role of font size and font style in younger and older adults' predicted and actual recall performance. *Aging, Neuropsychology, and Cognition*, *23*(3), 366–388. https://doi.org/10.1080/13825585.2015.1102194

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press.

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, *46*(2), 168–175. https://doi.org/10.1177/0098628319834381

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Rivers, M. L., Dunlosky, J., & Persky, A. M. (2020). Measuring metacognitive knowledge, monitoring, and control in the pharmacy classroom and experiential settings. *American Journal of Pharmaceutical Education*, *84*(5), 7730. https://doi.org/10.5688/ajpe7730

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, *41*(7), 1000–1011. https://doi.org/10.3758/s13421-013-0323-8

Tauber, S. K., Dunlosky, J., Urry, H. L., & Opitz, P. C. (2017). The effects of emotion on younger and older adults' monitoring of learning. *Aging, Neuropsychology, and Cognition*, *24*(5), 555–574. https://doi.org/10.1080/13825585.2016.1227423

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-paced study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024–1037. https://doi.org/10.1037/0278-7393.25.4.1024

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, *73*(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, *46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition*, *45*(4), 639–650. https://doi.org/10.3758/s13421-016-0681-0

Witherby, A. E., Tauber, S. K., & Dunlosky, J. (2021). Why do emotional stimuli influence judgments of learning? Theory, evidence, and future directions. *Trends and Prospects in Metacognition Research Across the Life Span*, 3–24. https://doi.org/10.1007/978-3-030-51673-4_1

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, *99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

Yin, Y., Shanks, D. R., Li, B., Fan, T., Hu, X., Yang, C., & Luo, L. (2023). The effects of emotion on judgments of learning and memory: A meta-analytic review. *Metacognition and Learning*, 1–23.

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

LIST 1	LIST 2	LIST 3	LIST 4
suicide	burn	spice	bug
winter	chief	city	hobby
seat	straw	salad	book
death	robbery	cream	aunt
wagon	basket	grave	wasp
killer	wound	prison	jail
hurt	crash	cut	brook
fence	junk	mist	blood
triangle	pilot	surgery	bullet
liquid	cancer	wrinkle	movie
monster	nest	hymn	trash
ache	dime	cloth	bomb
harp	zombie	tsunami	rat
stick	statue	flood	vomit
war	mirror	oven	shotgun
devil	shoe	steak	snow
valve	tyrant	mail	grill
sponge	bulb	fat	fairy
demon	cabin	poison	weapon
rape	burial	train	camp
disease	bench	riot	lady
virus	drown	tornado	poem
waist	stab	tent	peach
truck	rotten	thief	slap