The influences of valence and arousal on judgments of learning and on recall

Kathleen L. Hourihan[1] · Scott H. Fraundorf[2] · Aaron S. Benjamin[3]

Abstract Much is known about how the emotional content of words affects memory for those words, but only recently have researchers begun to investigate whether emotional content influences metamemory—that is, learners' assessments of what is or is not memorable. The present study replicated recent work demonstrating that judgments of learning (JOLs) do indeed reflect the superior memorability of words with emotional content. We further contrasted two hypotheses regarding this effect: a physiological account in which emotional words are judged to be more memorable because of their arousing properties, versus a cognitive account in which emotional words are judged to be more memorable because of their cognitive distinctiveness. Two results supported the latter account. First, both normed arousal (Exp. 1) and normed valence (Exp. 2) independently influenced JOLs, even though only an effect of arousal would be expected under a physiological account. Second, emotional content no longer influenced JOLs in a design (Exp. 3) that reduced the primary distinctiveness of emotional words by using a single list of words in which normed valence and arousal were varied continuously. These results suggest that the metamnemonic benefit of emotional words likely stems from cognitive factors.

✉ Kathleen L. Hourihan
   khourihan@mun.ca

1   Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X9, Canada

2   Learning Research and Development Center and Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

3   Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Emotion is a fundamental aspect of the human experience, and understanding how emotion influences our cognitive processes is an essential component of understanding cognition in general (Mandler, 1975). Emotional effects on memory could include memory for an emotional event, memory for neutral stimuli encountered while in an emotional state, and memory for emotional content encountered in a neutral state. Here, we consider the lattermost of these scenarios: how the emotional content of material encountered in an otherwise neutral state influences subsequent memory and metamemory.

We frequently encounter situations in which we learn and remember emotional information, such as when we read a news article recounting a natural disaster or celebration, or when we encounter emotionally evocative literature. Generally, emotional information is remembered better than neutral information (see Kensinger, 2009; Levine & Edelstein, 2009, for reviews); this has been observed with both pictures (e.g., Bradley, Greenwald, Petry, & Lang, 1992) and words (e.g., Kensinger & Corkin, 2003). In the present study, we examined whether and how emotion influences our *predictions* of future recall, and what these findings imply about the processes by which emotion influences memory and metamemory.

Metamemory for emotional information

Despite the vast amount of research examining memory for emotional information, relatively little research has focused on how *metamemory* (i.e., the monitoring and control of memory) is influenced by emotion. Some general evidence (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010), detailed

Springer

below, suggests that learners' predictions of recall can be sensitive to the emotional content of the information, but the origin of this relationship is not clear.

Some plausible hypotheses are suggested by the more general literature on metamemory. Extensive research in metamemory has examined *judgments of learning* (JOLs; Arbuckle & Cuddy, 1969): predictions of future memory performance elicited sometime after study but prior to test. Koriat's cue-utilization theory (1997) describes three types of cues that may be accounted for in making JOLs. *Intrinsic* cues are inherent to the study item itself, such as whether a word is concrete or abstract. *Extrinsic* cues are aspects of the study context not directly associated with the target items, such as repetition during presentation. JOLs made on these bases would reflect conscious, theory-based reasoning. By contrast, *mnemonic* cues are internal, subjective feelings about how well an item has been learned; they are less consciously identifiable, and rely more on a feeling of fluency (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin & Bjork, 1996). JOLs are generally quite sensitive to intrinsic cues, such as word frequency (Benjamin, 2003) and imageability (Begg et al., 1989), whereas they often fail to account for extrinsic cues (e.g., Koriat, 1997; Koriat, Bjork, Sheffer, & Bar, 2004). JOLs for emotional words may therefore be influenced by both conscious, theory-based (intrinsic) and unconscious, experience-based (mnemonic) contributions.

This literature suggests several different *potential* mechanisms by which emotion could influence JOLs. First, because the emotional content of a word is intrinsic rather than extrinsic, it would be likely to influence JOLs. Participants might explicitly believe that emotional information will be remembered better than neutral information, and when they encounter items that are obviously emotional (especially in the context of other items that are neutral), they purposefully assign higher JOLs to emotional items than to neutral items. That is, the effect of emotion on JOLs may reflect a conscious strategy that incorporates a salient intrinsic cue. We note that this possibility closely resembles a theory-based influence of emotion on JOLs (Koriat, 1997), although we here term this the *cognitive* account of emotion on JOLs to distinguish this hypothesis (about emotion effects on metamemory) from more general claims about potential influences on JOLs.

But, the above literature also implies that a second route by which emotion may influence metamemory is an unconscious, feeling-of-fluency effect (e.g., Benjamin & Bjork, 1996; Benjamin, Bjork, & Schwartz, 1998; Koriat, 1997). The physiological response associated with reading an emotional word may be perceived as fluency of processing and lead to predictions of greater recall (paralleling claims about emotion influences on *actual* recall; Scherer, 2005). We term this the *physiological* account[1] when applied to emotion influences on metacognition, but note that it is analogous to the more general experience-based influences on JOLs described by Koriat (1997). Supporting the physiological account, there is some general evidence that emotional words elicit physiological responses that would be relevant to mnemonic and metamnemonic processes: Encoding of emotional words is associated with greater activation in the amygdala than encoding of neutral words (e.g., Kensinger & Corkin, 2004), and emotional arousal enhances memory consolidation (e.g., Cahill & McGaugh, 1998). Also, in recognition memory, high-arousal stimuli lead to increased reports of recollection, as compared to low-arousal stimuli (e.g., Ochsner, 2000).

Thus, it is plausible that learners may make metamnemonic judgments about emotional information either on a conscious basis (a cognitive account) or an unconscious basis (a physiological account)—or both. In the present study, we sought to distinguish these possibilities.

Prior investigations of emotion and metamemory

Although the specific mechanisms by which emotion influences metamemory are unclear, the evidence to date is broadly consistent with the claim that JOLs are sensitive to emotional content. Zimmerman and Kelley (2010) compared negative and positive emotional words to neutral words. JOLs were consistently higher for both negative and positive words relative to neutral words. However, this pattern of JOLs was not entirely predictive of actual recall: Only on free recall tests were emotional words were consistently recalled better than neutral words; in cued recall, only positive words (not negative words) elicited higher recall than neutral words. Zimmerman and Kelley concluded that emotional content (regardless of whether it is positive or negative) is judged to make words more memorable; this strategy is often helpful but fails to account for other aspects of the encoding episode, such as the test format or the relatedness of word pairs, which may be more influential.

Zimmerman and Kelley's (2010) free recall findings were subsequently replicated by Tauber and Dunlosky (2012), who also included older adult participants in their study. Older adults' predictions about recall of emotional words are particularly interesting because older adults are more likely than young adults to show a memory benefit for positive emotional information over negative emotional information (e.g., Tomaszczyk, Fernandes, & MacLeod, 2008). However,

---

[1] The term *physiological* is used here in reference to prior work substantiating the claim that high-arousal words lead to a different physiological response; we did not directly measure physiological responses in the present study.

Tauber and Dunlosky ([2012](#)) found that although both groups recalled negative and positive words better than neutral words, older adults' JOLs did not show any particular bias for positive over negative words: Both young and older adults gave higher JOLs to negative words than to neutral words, but only young adults gave higher JOLs to positive words than neutral words. (We revisit these discrepancies between JOLs and actual recall in the General Discussion.)

Although these two studies stand as the first demonstrations that emotional content influences JOLs, they provide limited guidance on exactly *why* emotional content influences JOLs. Participants appeared to be using a heuristic to assign higher JOLs to emotional words than to neutral words, but that heuristic may have been based on an explicit belief, on an interpretation of a physiological response, or on some combination of belief and feeling. Our goal in the present study was to adjudicate among these possibilities.

Dimensions of emotionality

How can we distinguish cognitive (conscious) and physiological (unconscious) influences of emotion on metamemory? One strategy is to use more finely tuned manipulations of emotional content. Emotional stimuli differ from neutral ones on several dimensions, including both valence and arousal.[2] *Valence* indicates how pleasant a word is (with low numbers indicating negativity and high numbers indicating positivity; neutral items fall in between), and *arousal* indicates the energy level associated with a word. These two factors are the primary factors used by researchers to characterize emotional words and to differentiate lists of emotional words from lists of neutral words. However, with notable exceptions in the work of Kensinger and colleagues (e.g., Kensinger & Corkin, [2003](#), [2004](#)), they have frequently been confounded in memory and metamemory research: Emotional words in many studies differ from neutral words in having both more extreme valence (i.e., more negative or more positive) *and* higher arousal (e.g., Siddiqui & Unsworth, [2011](#); Talmi & Moscovitch, [2004](#)), including in the metamemory studies by Zimmerman and Kelley ([2010](#)) and Tauber and Dunlosky ([2012](#)).

The confounding of arousal and valence means that it is unclear which factor primarily drives the effects of emotion on metamemory. However, the two hypotheses we reviewed above make competing predictions about which factor (arousal or valence) drives effects of emotion on metamemory. If the

influence of emotion on metamemory occurs mainly via physiological arousal, then it is primarily arousal, not valence, that should influence metamemory. By contrast, if emotional influences on metamemory reflect the explicit application of naïve theory-driven strategies (i.e., responding to the relative salience of a class of emotional words), then *both* valence and arousal could influence metamnemonic judgments as long as they are sufficient to distinguish a relatively salient class of emotional words against a background of comparatively neutral words.

List composition and distinctiveness

A second way to test the mechanisms by which emotional content influences metamemory is by manipulating the list structure. As we noted above, one reason that emotional words might receive higher JOLs is that participants apply a strategy of assigning higher JOLs to words that are noticed to be intrinsically emotional, on the basis of a belief that emotional information should be remembered better than neutral information. But for participants to be able to apply a naïve theory of the effects of emotional content of memory, the memoranda must reveal to the participants their membership or nonmembership in a category of "emotional" items. Thus, list composition should be an important variable in determining the effects of emotional content on metamemory. This possibility for *metamemory* is suggested by prior results from *memory*: The memory benefit for emotional items occurs primarily in mixed-list designs, in which a single list contains discrete categories both of emotional and of neutral items, and not in pure-list designs (e.g., Dewhurst & Parry, [2000](#)). Similarly, the apparent benefit in memory for emotional information is eliminated if emotionally neutral words that are semantically related to one another are used as comparison stimuli (e.g., Talmi & Moscovitch, [2004](#)). These results suggest that emotional words require the background of neutral words to be present during learning in order to stand out as distinctive and show differences in memory. (This *primary distinctiveness*, based on context, can be contrasted with *secondary distinctiveness*, in which items stand out regardless of context due to their inherent unusualness; Schmidt, [1991](#)). If the same is true for emotional influences on *metamemory*, emotion should not affect JOLs when the study context does not make emotional words comparatively distinct.

By contrast, if the physiological process of encountering emotional information itself influences metamnemonic judgments, there is no reason to think that list composition should influence whether we observe emotion effects on metamemory. In all cases, learners are encountering and processing emotional information, and the physiological response should be similar or the same. Thus, JOLs for emotional

---

[2] The Affective Norms for English Words (ANEW) database (Bradley & Lang, [2010](#)) contains an additional dimension, *dominance*, which refers to being controlled versus being in control. In the ANEW database, dominance ratings correlate highly with valence ratings ($r = .83$; i.e., being controlled is negative, whereas being in control is positive). Most researchers do not address this factor; we also do not consider dominance in the present study.

words would be higher than for neutral words, regardless of whether the list composition provides a distinctive context.

These competing predictions have not been directly tested. The few studies on metamemory for emotional words (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010) have all used mixed-list designs in which the lists were composed of discrete categories of emotional and neutral words, which may have highlighted the emotional words by contrast to the neutral words. One set of studies—those by Tauber and Dunlosky (2012)—did vary the number of categories across experiments (positive, neutral, and negative versus only positive and neutral or only negative and neutral) and found no effect of this variable on JOLs. However, those studies only varied the *number* of discrete categories and not *whether* such discrete categories existed within the list at all.

Present experiments

The physiological and cognitive accounts make distinct predictions. According to a physiological account, emotion should influence JOLs primarily via arousal, rather than valence. Furthermore, list composition should have little influence on the emotional JOL effect because the effect directly results from the physiological experience associated with reading a given word. By contrast, under a cognitive account, JOLs may be influenced by emotion much more consciously. This account relies on two components: (a) participants have an explicit belief that emotional information is more memorable than neutral information, and (b) emotional content must be sufficiently salient at encoding, such that participants could explicitly categorize words as *emotional* versus *neutral*. According to this account, JOLs may be influenced by either valence or arousal, so long as a distinct set of emotional items is saliently different from a set of neutral items studied in the same context. This account would also predict that a study context that eliminates the distinctiveness of emotional items would therefore reduce or eliminate the salience of emotion, eliminating any influence of emotion on JOLs.

In the present study, we tested these competing predictions in two ways. First, in Experiments 1 and 2, we separately examined how JOLs are affected by arousal and by valence while holding the other property constant. Second, we examined these relations in two different types of mixed lists. In what we term a *discrete-levels* design (Exps. 1 and 2), each stimulus belongs to one of two distinct categories that together comprise the mixed list (e.g., a list that consists half of high-arousal items and half of low-arousal items). The sharp distinction between the emotional and nonemotional categories in these lists should make the emotional nature of the stimuli relatively salient and encourage the use of an explicit belief about the memorability of emotional stimuli. By contrast, in what we term a *continuous-levels* design (Exp. 3), the items

within the mixed list represent the full range of variance (e.g., some items are very high in arousal, some moderately high, some neutral in arousal, some moderately low, and some very low), with no categorical distinction between emotional and nonemotional items. These lists should make the emotionality of the stimuli less salient, and consequently should provide less or no support for the application of any explicit beliefs about the memorability of emotional stimuli.

Experiment 1

The goal of the first experiment was to investigate whether the emotional effect on JOLs is caused by an implicit, heuristic interpretation of the feeling that is experienced when reading a high-arousal, emotional word. Although past research has consistently shown that higher-arousal words receive higher JOLs (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010), those studies used high-arousal words that also differed in valence (whether positive or negative) from the low-arousal, neutral-valence comparison words. Thus, it is not clear whether arousal per se is responsible for the higher JOLs assigned to emotional words.

In Experiment 1, we used a discrete-levels manipulation of arousal while holding valence constant. Participants studied a list of neutral-valence words, half of which were low-arousal and half of which were high-arousal. Participants were asked to provide immediate JOLs for each word. If the influence of emotion on JOLs is driven by the emotional arousal of reading an emotional word, as in the physiological account, high-arousal words should receive higher JOLs than low-arousal words. (Note that such a result would also not necessarily be incompatible with the cognitive account, because participants might identify the high-arousal words as a distinctive category.) Alternatively, high-arousal words might not receive higher JOLs than low-arousal words once they were equated in valence; such a result would provide strong evidence against the physiological account.

Method

Participants Forty undergraduate students at the University of Illinois at Urbana-Champaign participated for course credit.

Materials Two lists of 20 words were selected from the ANEW database (Bradley & Lang, 2010; see Appendix A for the word lists). The top half of Table 1 displays the mean valence and arousal ratings and word frequencies for the two Experiment 1 lists. The lists differed significantly on arousal [$t(38) = 33.48$, $p < .001$], but not on valence [$t(38) = 0.18$, $p = .860$]. As a measure of word frequency, we used the SUBTLEXus log word frequency, the frequency measure that best predicts human behavior in a variety of common

Table 1 Characteristics of the word lists in Experiments 1 and 2

	Valence (1–9)		Arousal (1–9)		Word Frequency (number per million)	
	Range	Mean	Range	Mean	Range	Mean
Experiment 1						
High arousal	[4.02, 5.93]	4.96 (0.57)	[6.25, 7.93]	6.63 (0.42)	[1.28, 3.78]	2.47 (0.69)
Low arousal	[4.10, 5.58]	4.93 (0.39)	[2.65, 3.36]	3.10 (0.21)	[1.99, 4.14]	2.80 (0.54)
Experiment 2						
Negative valence	[2.06, 2.98]	2.67 (0.24)	[4.53, 5.42]	5.07 (0.26)	[1.42, 3.56]	2.37 (0.60)
Neutral valence	[4.50, 5.43]	4.92 (0.24)	[2.90, 7.36]	4.89 (0.80)	[0.60, 3.48]	2.36 (0.79)

Standard deviations are displayed in parentheses below their respective means

psycholinguistic tasks (Brysbaert & New, 2009). The two lists did not differ significantly in log word frequency [$t(38) = 1.717$, $p = .094$], and the direction in which they did differ runs counter to the expected benefit for high-arousal words (which were, on average, less frequent).

Procedure The experiment consisted of a study phase, a distractor phase, and a test phase. In the study phase, participants were informed they would be shown a list of words to study, presented one at a time, and would later be asked to recall as many of the words as they could. All 40 words were presented, one at a time, in random order. Each trial began with a 500-ms blank screen. The word then appeared at the center of the screen in 18-point Arial font (black on a white background) for 2,000 ms. After another 500-ms blank screen, participants were asked to make a prediction of how likely they thought they would be to recall the word (i.e., a JOL). They provided a response from 1 (*sure I will NOT remember*) to 6 (*sure I WILL remember*) by pressing the appropriate key on the keyboard; this scale (with labels) was presented on the bottom of the screen. All JOLs were self-paced. An additional 500-ms blank screen preceded the next trial.

After the final study trial, participants completed a distractor task that consisted of labeling state map outlines for 5 min. (Piloting showed that participants were unlikely to complete all 50 states in this time; if participants did complete all 50, the state maps were re-presented in a new random order.) They were then given instructions for the test phase. Participants were asked to recall as many of the words from the study phase of the experiment that they could remember. They were instructed to type the words one at a time and to press ENTER to submit each word. There was no time limit, and participants were instructed to guess when uncertain.

Results and discussion

The top half of Table 2 displays the mean JOLs and proportions of recall in Experiment 1. Participants predicted better

recall of high-arousal than of low-arousal words, $t(39) = 5.23$, $p < .001$, $d = 0.67$, 95 % confidence interval (CI) of difference [0.24, 0.55]. However, this prediction was inaccurate, since recall did not differ significantly as a function of arousal, $t(39) = 0.21$, $p = .836$, $d = 0.03$, 95 % CI of difference [–.04, .05]. The number of intrusions was low and thus was not further analyzed ($M = 0.88$, $SD = 1.42$).

Metamnemonic resolution (i.e., the degree to which an individual could predict which particular items were more or less likely to be recalled) was computed using the signal-detection-based d_a measure (Benjamin & Diaz, 2008; Green & Swets, 1966; Masson & Rotello, 2009), and is also displayed in Table 2. When applied to metamnemonic accuracy, d_a is essentially a measure of the degree to which participants can discriminate, at the time of the JOL, between items that will later be remembered and items that will later be forgotten. The accuracy of the JOLs predicting recall did not differ significantly between high- and low-arousal words, $t(39) = 1.40$, $p = .169$, $d = 0.24$, 95 % CI of difference [–0.38, 0.07]. That is, arousal had no significant influence on *relative* metamnemonic accuracy; higher arousal does not appear to convey better information about which individual items are more or less likely to be recalled.

To summarize, even though arousal did not affect actual recall, participants predicted that high- arousal words would

Table 2 Mean memory and metamemory performance in Experiments 1 and 2

	JOL (1–6)	Recall (proportion)	Resolution (d_a)
Experiment 1			
High arousal	3.59 (0.10)	.30 (.02)	0.62 (0.10)
Low arousal	3.19 (0.08)	.31 (.03)	0.47 (0.10)
Experiment 2			
Negative valence	3.73 (0.10)	.29 (.02)	0.53 (0.06)
Neutral valence	3.20 (0.12)	.22 (.02)	0.39 (0.10)

Standard errors of the means are displayed in parentheses below their respective means.

be more likely to be recalled. Furthermore, this pattern of higher JOLs to high-arousal than to low-arousal words emerged even though valence was held constant. This result could provide some support for a physiological account of the effects of emotion on JOLs: The high-arousal words may have evoked a different physiological response when read (Hamann, 2001; Kensinger & Corkin, 2004), and this response was interpreted as being indicative of increased future recall, similar to the mechanism by which people believe that highly emotional events will be remembered better (even though they are not; Talarico & Rubin, 2003). However, these results do not rule out a cognitive account; it is also possible that participants consciously interpreted the high-arousal words as being more "emotional" and explicitly assigned higher JOLs to emotional words.

Furthermore, the word lists in Experiment 1 may in some senses not have been fully equated in valence. Although the mean valences did not differ across lists, the standard deviation of the valence ratings was somewhat higher for high-arousal than for low-arousal words. That is, the high-arousal word list may have included more words of "ambivalent" valence rather than of truly neutral valence (e.g., *pregnant*, which may be perceived as positive by some individuals, but negative by others).

A more stringent test of the two accounts of the effects of emotion on metamemory, and one independent of the "ambivalent"-valence concern, is the converse question: Does valence affect JOLs in the absence of differences in arousal? A strong prediction of the physiological-arousal account is that emotional words should be judged no differently than neutral words when arousal is held constant. Thus, in our second experiment, we examined whether negative emotional words would still receive higher JOLs than neutral words when there were no differences in arousal.

Experiment 2

In Experiment 2, we used a discrete-levels manipulation of valence, in which participants studied a list that included both negative-valence and neutral-valence words. The word lists were selected to differ only in valence; arousal was held constant at approximately the midpoint of the scale so that words were neither particularly low nor high in arousal. Recall has been shown to be affected by valence independently of arousal (Kensinger & Corkin, 2003). Given that the two word types were presented in a mixed-list design, which should make the negative words distinctive relative to the neutral words (Talmi, 2013), we predicted that the negative-valence words would be recalled better than the neutral words.

Our primary interest, however, was metamemory. The cognitive and physiological accounts described above make competing predictions about whether valence influences JOLs in

the absence of differences in arousal. If emotion influences JOLs only through physiological arousal, then we should not see higher JOLs for our negative-valence words because they are equal in arousal to our neutral-valence words. By contrast, if emotion influences JOLs through participants' cognitive assessment of the intrinsic word properties, then JOLs should be higher for negative-valence than for neutral-valence words: Valence is a relatively salient intrinsic cue that will likely lead to higher JOLs for negative-valence words when they appear in a study context that includes a separate, distinct category of neutral-valence words.

Method

Participants Forty undergraduate students at the University of Illinois at Urbana-Champaign participated for course credit. None had participated in Experiment 1.

Materials Two lists of 20 words (see Appendix B) were selected from the ANEW database. The bottom half of Table 1 presents the mean valence and arousal ratings for the two Experiment 2 word lists.[3] The two lists differed significantly on valence [$t(38) = 28.98$, $p < .001$], but not arousal [$t(38) = 0.92$, $p = .365$]. Table 1 shows the mean word frequencies (occurrences per million) for the two lists; log word frequencies also were not significantly different between the two lists, $t(38) = 0.045$, $p = .964$.

Procedure The procedure was identical to that of Experiment 1.

Results and discussion

The bottom half of Table 2 displays the mean JOLs and proportions of recall in Experiment 2. Participants predicted better recall of negative-valence than of neutral-valence words, $t(39) = 6.27$, $p < .001$, $d = 0.74$, 95 % CI of difference [.36, .70]. This prediction was accurate: Recall was also significantly higher for negative than for neutral words, $t(39) = 3.89$, $p < .001$, $d = 0.58$, 95 % CI of difference [.03, .11]. The number of intrusions was again low ($M = 1.00$, $SD = 1.43$) and was not analyzed any further. Metamnemonic resolution (d_a) is displayed in Table 1. JOLs predicted free recall equivalently for the negative and neutral words, $t(39) = 1.24$, $p = .222$, $d = 0.27$, 95 % CI of difference [–.37, .09].

Replicating Kensinger and Corkin (2003), negative-valence words were more likely to be recalled than neutral-

---

[3] The range of arousal values for the neutral-valence word list appears high due to the inclusion of a single word with a high arousal rating ("hysterical"). We also analyzed the results of Experiment 2 while excluding this item from the data, and the results did not change from what is presented here. Therefore we report the full results, including all items from both word lists.

valence words, even when the words were equivalent in terms of arousal. The novel finding in this experiment was the metamnemonic one: JOLs, like recall, were also sensitive to differences in valence; specifically, negative words were (accurately) predicted to be more likely to be recalled than neutral-valence words. Because this difference emerged in spite of the words being equated on arousal, this result provides evidence against the claim that emotional influences on memory and metamemory reflect the physiological arousal experienced by learners. Rather, participants' sensitivity to the influence of valence on subsequent recall is more consistent with the idea that the effects of emotional content on metamemory are mediated at least partly through a cognitive evaluation of the intrinsic properties of the stimuli. The negative-valence words may have been explicitly noticed as being emotional when compared to the background of neutral-valence words in the same list; as a result, participants could consciously assign higher JOLs to the negative words on the basis of a belief that emotional information should benefit recall. (This difference in valence, however, had no influence on the relative ability to predict which individual words would be recalled and which would not.)

By comparison, Experiment 1 showed that arousal also affected JOLs, even when valence was held constant. That effect was seemingly consistent with a physiological account in which learners implicitly interpret feelings of arousal as being predictive of future recall. How can the results of these two experiments be reconciled? One interpretation is that *both* cognitive and physiological factors influence JOLs for emotional words, with the effects in Experiment 1 reflecting physiological arousal and those in Experiment 2 reflecting cognitive factors. However, another, more parsimonious account is that the effects in both experiments reflect cognitive mechanisms alone. The results of Experiment 2 strongly suggest that the source of emotion effects on JOLs is at least partially cognitive because (contrary to a physiological account) participants provided higher JOLs to negative words, even though there was no difference in arousal. It is also possible that cognitive mechanisms could explain the results of Experiment 1, if the high-arousal words in Experiment 1 were consciously interpreted as being more emotional than the low-arousal words. Note, however, that the results of these two experiments do not necessarily mean that arousal can *never* influence JOLs through implicit interpretation of a physiological response. For instance, it is possible that cognitive and physiological factors have an interactive influence, such that implicit arousal effects combine with explicit valence effects to produce higher JOLs for emotional words (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010).

To more clearly determine whether physiological or cognitive factors primarily influence the JOLs for emotional words, in Experiment 3 we used the list structure to reduce or eliminate the influence that cognitive factors could potentially exert

on recall prediction. Experiment 3 also addressed the limitation that, to this point, we have considered only certain ranges and combinations of arousal and valence. That is, in Experiment 1 we considered only high versus low arousal, while valence was held constant at a neutral level. We did not consider low- and high-arousal words of negative or positive valence, nor any moderate-arousal words at any of these levels of valence; our results thus may not generalize outside the ranges examined. The same logic applies to Experiment 2: We compared negative- and natural-valence words of moderate arousal, but not positive-valence words, nor did we compare neutral- to negative-valence words when the words were all high or all low in arousal (or include positive valence at high or low arousal levels). Thus, in addition to examining the role of list structure, we also sought in Experiment 3 to test whether the two emotional factors—valence and arousal—influence JOLs (and recall) independently and/or interactively.

Experiment 3

The cognitive account makes a key prediction about how list structure should influence JOLs for emotional words. If emotional words typically receive higher JOLs because they are distinctive, the effect of emotion on JOLs should be eliminated if that emotion information is no longer distinctive (as appears to be the case for emotion effects on recall itself; Talmi, 2013). That is, emotional words may normally stand out as distinctive in comparison to a weaker background of neutral words, and therefore may lead to higher JOLs (and recall) due to primary distinctiveness (Schmidt, 1991), but not if there is not a clear background of other word types.

To test this prediction, instead of defining discrete "emotional" or "neutral" categories, we created a larger pool of words selected from the full ranges of both arousal and valence, and used a continuous-levels design. That is, we selected items from a range of arousal values, ranging from low to high, and also from a range of valence values, ranging from low to high, for inclusion in the experiment. (See Appendix C for the full word pool.) Each participant studied a series of lists that were composed of a random selection of words from the pool; because these words varied continuously across the full range of arousal and valence values, emotional words no longer formed a distinct class of words, eliminating the primary distinctiveness present in the first two experiments. As in the earlier experiments, participants provided an immediate JOL for each word during study. Using the valence and arousal values from the ANEW norms for each word, we then used linear mixed-effects regression to determine how valence and arousal contributed to the predicted and actual recall of emotional words.

The physiological account predicts that arousal should influence JOLs even in this continuous-levels design. Reading words with higher arousal levels should lead to the interpretation that the words should be memorable, and arousal should thus increase JOLs. The predictions of the physiological account for the effects of valence are less clear. Valence may not have an independent influence on JOLs, given that differences in valence may not produce different physiological responses when changes in arousal are absent. However, other researchers have demonstrated that high arousal and negative valence *together* had a superadditive effect on recall (as compared to neutral-valence, low-arousal words; Kensinger & Corkin, 2003). Arousal may therefore *interact* with valence to influence recall, such that both low valence (negative) and high valence (positive) increase recall, but only when arousal is also high. It is possible that a similar interaction might be observed in JOLs.

By contrast, a cognitive account of the effects of emotion on metamemory would predict that neither valence nor arousal should influence JOLs in Experiment 3. Under this account, participants must explicitly notice that discrete classes of items—emotional and neutral—are present in the study context and assign JOLs in accordance with their conscious beliefs about the influence of emotion on memory. Although emotional items can benefit from primary distinctiveness when they are presented against a background of neutral words, the continuous-levels design used in Experiment 3 should drastically reduce (or eliminate) the salience of the words' emotional content. Thus, without a clear contrast between item classes, participants would be unlikely to apply any belief that emotional items would be more memorable than neutral items when providing JOLs.

Method

Participants Eighty undergraduate students from Memorial University of Newfoundland participated in exchange for either course credit or payment.

Materials Eighty words[4] were selected from the ANEW database. (See Appendix C for the word list.) Words were selected in order to obtain approximately equal numbers of words from the bottom third, middle third, and top third of both the valence and arousal ratings. Overall, valence ranged from 1.55 to 8.72 ($M = 5.22$, $SD = 2.06$), and arousal ranged from 2.50 to 8.17 ($M = 5.12$, $SD = 1.46$). The log word frequency ranged from 0.29 to 590.69 ($M = 38.63$, $SD = 80.12$), representing from 1.20 to 4.48 occurrences per million words ($M = 2.77$, $SD = 0.71$).

---

[4] For one word, *tatter*, we discovered that no SUBTLEXus word frequency was available, so we excluded this item from the analysis.

Procedure Experiment 3 was similar to the first two experiments, but it contained four study–test cycles. This procedural change was introduced to avoid floor effects in free recall due to the increased number of words. Participants were informed that they would be asked to study and recall several different lists of words. Each study cycle presented 20 words randomly selected from the pool of 80, with the constraint that no word was used in more than one cycle; thus, all participants eventually studied the same 80 words, but they encountered different random list compositions. All study trial timings were identical to those in the first two experiments. The distractor intervening between study and test was now a mental rotation task, in which participants selected which of two shapes was a rotated version of a target shape, and the distractor-task duration was reduced to 3 min. The test procedure was identical to those of the first two experiments. After participants indicated that they had recalled as many words as they could, they completed simple arithmetic problems for 1 min prior to the next study phase.

Analytic strategy Several aspects of Experiment 3 called for a different analytic approach. First, the predictors of interest varied continuously across the full range of arousal and valence, rather than being presented in discrete categories. Although it would be possible to collapse such variability into a smaller number of categories (e.g., with a median split) for a factorial analysis of variance (ANOVA), such techniques would greatly reduce statistical power (Cohen, 1983). Second, we now had multiple predictors of interest (arousal, variance, and word frequency) and were interested in assessing the effect of each while holding the others constant.

Both of these problems can be solved by linear mixed-effects regression (Baayen, Davidson, & Bates, 2008; Jaeger, 2008; for applications to memory and metamemory, see Fraundorf, Benjamin, & Watson, 2013; Fraundorf, Watson, & Benjamin, 2010; Freeman, Heathcote, Chalmers, & Hockley, 2010; Hourihan, Fraundorf, & Benjamin, 2013; Murayama, Sakaki, Yan, & Smith, 2014). Like all multiple-regression models, these models can incorporate multiple predictors of interest (termed *fixed effects* in the mixed-effects regression context), including continuously varying quantities, such as valence. Linear mixed-effects regression can also control for variability across multiple *random effects*—effects for which the observed categories are sampled out of a larger population; in the present context, this included both subjects (sampled out of a population of possible participants) and items (sampled out of a pool of possible words).

We fit three models corresponding to the same aspects of behavior we had analyzed in the first two experiments. In the first, most crucial model, we examined metamemory by analyzing participants' JOLs as a function of word frequency, arousal, valence, and the Arousal × Valence interaction. We considered two ways valence might affect JOLs. JOLs might

reflect a word's degree of positivity, with negative words being given the lowest JOLs, neutral words being given moderate JOLs, and positive words the highest JOLs. This specific possibility seems less likely, given that our second experiment and past research showed that negative words were given higher JOLs than neutral words (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010). A more likely alternative is that JOLs would reflect the strength of a word's valence, with both negative and positive words receiving higher JOLs than neutral words. To test both possibilities, we included both a linear polynomial contrast for valence (capturing the former possibility) and a quadratic polynomial contrast (capturing the latter). All predictors were mean-centered; centering variables produces estimates corresponding to ANOVA main effects (Cohen, Cohen, West, & Aiken, 2003).

The second model was a mixed-effects logit model (Jaeger, 2008) that modeled actual recall (specifically, the log odds of correctly recalling each item) as a function of the same word properties as above. Finally, we examined the relative accuracy of participants' metamnemonic predictions with another mixed-effects logit model that tested whether the JOL assigned to each item was associated with higher odds of recall (Murayama et al., 2014): A significant positive relationship between JOL and the odds of recall would mean that participants accurately assigned higher JOLs to words they were more likely to recall, whereas a null relationship would indicate that the JOLs were unrelated to the odds of recall.

In all mixed-effects regression models, the variance in an effect across subjects or items (e.g., variance across participants in how sensitive they are to word frequency) can be captured by random slope parameters. Following Barr, Levy, Scheepers, and Tily (2013), we included a conservative near-maximal random-effects structure that omitted only the correlations between random effects. All models were fit in the R environment for statistical computing, using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015).

To measure effect size in the mixed-effect regressions, we used the standardized regression coefficient, which measures the effect of a one-standard-deviation change in an independent variable in terms of standard deviations of change in the dependent variable.

Results

We excluded eight JOL trials on which participants pressed a key that was not a valid response, which affected less than 1 % of the data.

Metamemory We first examined which variables influenced participants' JOLs. Table 3 displays the parameter estimates from this model. Word frequency significantly predicted JOLs, with each one-unit increase in log word frequency increasing the average JOL by 0.27 (95 % CI: [.17, .37]), $t =$

5.63, $p < .001$. However, neither arousal, nor valence, nor their interaction significantly influenced JOLs.

Memory Next, we turned to which variables influenced free recall. Table 4 displays the parameter estimates[5] from the mixed-effects model of actual recall. Analogous to participants' predictions, only word frequency significantly influenced recall: Each one-unit increase in log word frequency increased the odds of recall by 1.25 times (95 % CI: [1.04, 1.50]), $z = 2.32$, $p < .05$.

The mean numbers of intruded words in free recall were relatively low across the four study–test cycles ($M = 0.79$, $SD = 0.98$; $M = 0.72$, $SD = 0.86$; $M = 0.80$, $SD = 1.10$; and $M = 1.16$, $SD = 1.55$). A repeated measures ANOVA indicated that intrusions significantly increased on the fourth test relative to the first three (all $ps \leq .05$); no other differences were significant.

Metamnemonic resolution Finally, we examined whether the JOL assigned to each item actually predicted the odds that it would later be recalled. The results from this model, in Table 5, confirmed that participants' JOLs indeed identified which items were relatively more likely to be recalled: For each one-unit increase on the JOL scale (which ranged from 1 to 6), the odds of subsequent recall increased 1.47 times (95 % CI: [1.38, 1.57]), $z = 11.34$, $p < .001$.

Discussion

Neither valence nor arousal had any significant influence on the JOLs in Experiment 3. This result contrasts with those obtained in mixed-list designs with discrete *emotional* versus *neutral* categories within the lists (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010; as well as our first two experiments); in those designs, emotional words received higher JOLs than neutral words, even when memory was not actually influenced by emotion in the way that participants predicted. One possible explanation for these prior findings was that participants used the physiological experience associated with reading an emotional word as a cue to predict that recall would be higher than for words that did not produce a similar response (Tauber & Dunlosky, 2012). This physiological explanation cannot provide a persuasive account of the present results. The participants in Experiment 3 certainly encountered words that were high-arousal and either positively or negatively valenced. If JOLs were directly based on physiological response, these emotional words should have received higher JOLs, yet neither valence nor arousal (nor their interaction) systematically influenced JOLs. The results are

---

[5] The raw parameter estimates for the logit models (presented in Table 3, Table 4, and Table 5) refer to the *log odds* of correct recall. To facilitate interpretation, the main text presents estimates that have been transformed to effects on the *odds*.

Table 3 Fixed-effect estimates for linear mixed-effects model of judgments of learning in Experiment 3

Fixed Effect	$\hat{\beta}$	SE	95 % CI	t	p	Standardized Coefficient
Intercept (mean rating)	3.56	0.08	[3.40, 3.72]	45.77	<.001	0.01
Arousal	−0.02	0.02	[−0.06, 0.02]	−1.09	.28	−0.02
Word frequency	0.27	0.05	[0.22, 0.32]	5.63	<.001	0.14
Valence: Linear effect	4.91	2.86	[−0.70, 10.52]	1.72	.09	3.43
Valence: Quadratic effect	3.88	2.49	[−1.00, 8.76]	1.56	.12	2.71
Arousal × Linear Valence	−0.68	1.59	[−3.80, 2.43]	−0.43	.67	−0.69
Arousal × Quadratic Valence	−0.66	1.57	[−3.74, 2.42]	−0.42	.67	−0.67

SE = standard error.

instead more consistent with the idea that cognitive factors best explain why emotional words receive higher JOLs. The composition of the Experiment 3 word lists did not clearly contrast emotional with neutral words, removing the cognitive influence of primary distinctiveness from both recall and JOLs, and apparently eliminating the emotion effect on JOLs. Thus, the fact that JOLs have previously been shown to be higher for emotional than for neutral words is likely to have arisen because participants explicitly noticed emotional words by their contrast to the neutral words in a mixed study list; only in those cases did participants consciously assign higher JOLs to emotional words, presumably on the basis of the belief that emotional information should be more memorable than neutral information (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010).

A similar method and similar conclusions have been reported to explain why emotional words are often recalled better than neutral words (Talmi, 2013). Although emotional words (typically selected to vary in both valence and arousal; e.g., Siddiqui & Unsworth, 2011; Talmi & Moscovitch, 2004; Zimmerman & Kelley, 2010; but see Kensinger & Corkin, 2003, 2004) are often recalled better than neutral words, such findings are usually observed with discrete-level versions of the mixed-list design, in which a study list contains only two or three types of items, and emotional words have the opportunity to stand out from the neutral

words (Talmi, 2013). When the influences of distinctiveness and organization are controlled at study by presenting pure lists of emotional and neutral words (rather than mixed lists), and/or when the neutral comparison words are semantically related, and thus themselves categorically related, the recall benefit for emotional words is eliminated (Dewhurst & Parry, 2000; Talmi & Moscovitch, 2004). The results of our Experiment 3 provide further evidence that superior recall of emotional information may reflect cognitive mechanisms, such as distinctiveness or relatedness, rather than the physiological effects of emotion per se (see also Talmi, 2013).

One possible counterexplanation is that the differences between Experiment 3 and the prior experiments are driven not by list composition per se, but rather by the fact that Experiment 3 (unlike Exp. 2) contained positively valenced words in addition to negatively valenced words. Perhaps qualitatively different mechanisms underlie metamnemonic judgments for positive versus negative items. However, there is evidence against this explanation. The linear effect of valence included in the statistical model specifically tested for an effect of positive versus negative valence; however, no such effect was observed. Rather, Experiment 3 revealed no effect of valence at all, whether positive or negative. Moreover, previous studies that had included both positive and negative emotional words in a discrete-levels design had shown that both valence categories led to similarly increased JOLs (as

Table 4 Fixed-effect estimates for linear mixed-effects model of recall in Experiment 3

Fixed Effect	$\hat{\beta}$	SE	95 % CI	Wald z	p	Standardized Coefficient
Intercept (mean log odds of recall)	−0.46	0.10	[−0.26, −0.66]	−4.40	<.001	−0.46
Arousal	−0.06	0.05	[−0.16, 0.04]	−1.42	.16	−0.10
Word frequency	0.22	0.09	[0.04, 0.40]	2.34	<.05	0.16
Valence: Linear effect	−0.80	5.08	[−10.76, 9.16]	−0.16	.88	−0.80
Valence: Quadratic effect	0.36	5.24	[−9.91, 10.63]	0.07	.94	0.36
Arousal × Linear Valence	−0.97	3.49	[−7.81, 5.87]	−0.28	.78	−1.13
Arousal × Quadratic Valence	−5.40	3.45	[−12.16, 1.36]	−1.57	.12	−7.89

SE = standard error

Table 5 Fixed-effect estimates for linear mixed-effects model of metamnemonic accuracy in Experiment 3

Fixed Effect	$\hat{\beta}$	SE	95 % CI	Wald z	p	Standardized Coefficient
Intercept (mean log odds of recall)	−0.50	0.10	[−0.30, −0.70]	−5.07	<.001	−0.50
JOL	0.39	0.03	[0.33, 0.45]	11.34	<.001	0.55

SE = standard error

compared to neutral-valence words), suggesting that both positive- and negative-valence items can lead to higher JOLs, given a distinctive list composition (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010).

Note that, although we found that neither valence nor arousal significantly influenced JOLs, inspection of the regression results revealed at least some numerical trends toward an effect of valence on JOLs and an Arousal × Valence interaction in recall. However, we are reluctant to overinterpret these patterns, given that they did not attain conventional levels of significance despite our comparatively large sample size. Furthermore, the reliable effects of word frequency on both JOLs and recall indicate that the linear mixed-effects regression was not simply insensitive to all influences on memory and metamemory; thus, the null effects of arousal and valence cannot simply be attributed to an insensitive statistical method.

Indeed, in the absence of emotion effects, word frequency was the only significant predictor of word recall and JOLs in Experiment 3. An effect of word frequency was expected, because it is well-known that high-frequency words are recalled better than low-frequency words (e.g., Hall, 1954).

General discussion

We conducted three experiments to examine how and why emotional content influences predictions of future memory. The first two experiments separately examined the influences of two different aspects of emotion—arousal and valence—on JOLs. Participants provided higher JOLs to words with emotional content. Crucially, emotional words received higher JOLs even if they differed from neutral words only in valence and not in arousal, suggesting that these metamemory effects do not solely reflect physiological emotional arousal. Our third experiment presented participants with word lists with continuously varying levels of arousal and valence, which minimized the primary distinctiveness associated with these variables. In this design, neither recall nor JOLs were reliably influenced by valence and/or arousal, further suggesting that emotion effects on metamemory and memory do not directly reflect physiological factors associated with emotion, but instead reflect cognitive factors at encoding, such as the relative distinctiveness of the items in a study list.

The mechanisms underlying emotion effects on metacognition

What accounts for the influence of emotion on metamemory? One possibility is that participants interpret their physiological emotional arousal as a feeling of fluency of processing and thus rate emotional words as more memorable. After all, such unconscious, mnemonic cues often have a strong influence on metamemory (Koriat, 1997). Experiment 1 did show that participants gave higher JOLs to high-arousal words than to low-arousal words even when all words were neutral in valence. However, given the results of our other two experiments and the fact that arousal did not actually influence recall, it seems unlikely that these JOL differences reflect an implicit interpretation of the physiological effects of arousal. Rather, a cognitive explanation is more likely: Higher JOLs for high-arousal words were driven by the contrast to the low-arousal words in the mixed-list design, coupled with participants' belief that these high-arousal experiences would be more memorable than lower-arousal experiences (Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010).

Indeed, Experiment 2 provided strong evidence against an explanation based purely on physiological arousal: Participants gave higher JOLs to negative-valence words than to neutral-valence words, even though both word types had equivalent, moderate levels of arousal. This result rules out a direct effect of physiological arousal and suggests that, instead, participants were likely consciously responding to valence on the basis of its distinctiveness in the discrete-levels, mixed-list context. That is, the negative words may have received higher JOLs than the neutral words because they stood out in contrast to the neutral words present in the same list (Talmi, 2013).

This explanation suggests that if the negative and neutral words did not form such distinct categories, JOLs would no longer differ between negative and neutral words, because there would no longer be differences in distinctiveness (Dewhurst & Parry, 2000). We tested this prediction in Experiment 3: Participants studied words that were high or low in arousal and also high or low in valence, but the words were studied in lists that did not enhance their distinctiveness by including a clear background of lower-arousal, neutral-valence words. Consistent with the predictions of the cognitive account, we no longer observed any effects of emotional

content on either JOLs or actual recall. These results further bolster the idea that JOLs for emotional words in fact reflect list composition rather than emotional factors per se. Had participants made JOLs to emotional words on the basis of an interpretation of the subjective physiological response associated with reading an emotional word, then we should have seen higher JOLs for emotional words than neutral words even in the randomly composed lists used in Experiment 3. We did not. Instead, it seems likely that the higher JOLs observed for emotional stimuli in other experiments arises because the typical list structure used more clearly contrasts emotional items with neutral items (as in Exps. 1 and 2, and in Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010), increasing the saliency of the emotional content of the words and leading participants to apply an explicit belief about the memorability of emotional words (and often to recall them better as well; Talmi, 2013).

However, as we described above in the discussion of Experiment 2, our results do not necessarily mean that arousal can *never* influence JOLs through implicit interpretation of a physiological response. We have demonstrated that arousal alone has a negligible influence on JOLs in the absence of a distinctive list structure that encourages theory-based use of arousal as a cue for predicting recall. But, especially in discrete-levels experimental designs, it is possible that both cognitive and physiological factors could operate interactively in leading participants to assign higher JOLs to emotional words. For example, considering our Experiment 1, studying words with differences in arousal may have led participants to consciously focus on the fact that the study list consisted of two different types of words; the requirement to provide JOLs for each word may have then led participants to consciously rely on this experience in providing JOLs (see Mueller, Dunlosky, Tauber, & Rhodes, 2014). Thus, fluency (or experience) may potentially be incorporated in a primarily cognitive (or theory-based) strategy. Indeed, future work could assess this possibility by explicitly querying participants (e.g., in a posttest survey) about whether they noticed the list composition. Another relevant method may be to vary emotional factors only between subjects or between lists in a pure-list design; our cognitive account predicts that emotion effect on JOLs should be absent in these cases because the emotion words would not have a background of neutral words to contrast with.

But, not all influences on metamemory need to be distinctive in order to have an effect. In Experiment 3, participants correctly predicted that higher-frequency words were more likely to be recalled than lower-frequency words even though the lists were not designed to clearly contrast low-frequency and high-frequency words. That is, unlike arousal and valence, it does not appear to be necessary for word frequency to be distinctive in order for participants to account for it when making JOLs. Word frequency might influence JOLs because,

in the absence of distinctive categories of words on which to base JOLs at encoding, participants seem to rely on familiarity (e.g., Begg et al., 1989; Benjamin, 2003). In the case of free recall testing (and especially in our Exp. 3), it is quite sensible to expect superior recall of higher-frequency (and thus more familiar) words relative to lower-frequency words.

An additional factor that may have influenced JOLs and recall is concreteness (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003). However, the ANEW database (Bradley & Lang, 2010) only includes norms of emotional factors, and we were only able to obtain a measure of concreteness for about half of our emotional word stimuli (Coltheart, 1981). Although our lists did not significantly differ in concreteness given the available norms, we cannot fully rule out the possibility that our lists were not equated on concreteness and that this may have influenced JOLs, recall, or both. Future studies should attempt to replicate our findings while controlling for the concreteness of the emotional word stimuli.

Metacognitive illusions

Although our primary interest was participants' metamemory (i.e., JOLs), we also assessed participants' actual recall. The conditions to which participants assigned higher JOLs were often those that actually produced superior recall (i.e., in Exps. 2 and 3), but not always. Specifically, participants in Experiment 1 gave higher JOLs to high-arousal (but neutrally valenced) words than neutral words, even though arousal had no influence on actual recall (see also Kensinger & Corkin, 2004, for a similar effect in recognition). Interestingly, if the recall test had taken place after a more substantial delay (i.e., long enough to allow for memory consolidation), JOLs may have more accurately reflected the pattern of recall, as arousal has been shown to increase memory consolidation (e.g., Cahill & McGaugh, 1998; see also Mather & Sutherland, 2011). However, this would likely be a coincidental increase in JOL accuracy, as participants have been shown to be poor at predicting the effects of retention intervals on memory (e.g., Koriat et al., 2004). At least for neutral-valence words, it would appear that high arousal can be added to the list of word qualities or encoding tasks that participants falsely believe will influence recall, such as large font size (Rhodes & Castel, 2008), loudness (Rhodes & Castel, 2009), and repeated vocal response (Castel, Rhodes, & Friedman, 2013).

This pattern, in combination with the apparent cognitive locus of emotion effects on JOLs, implies that some of the metacognitive illusions mentioned above may rely not only on differences in perceived fluency between conditions, but in the use of a distinctive list composition that makes increases the salience of those fluency differences. Considering the font size example, Rhodes and Castel (2008) argued that participants process words in large font more fluently than words in small font, leading to increased JOLs but not to increased recall.

The JOL effect persisted even when participants had experience with a recall test in which they did not recall large-font words any better than small-font words and even when participants were explicitly told that font size does not affect recall. The effect was only eliminated when all words (regardless of font size) were made difficult to process by using alternating letter case, leading Rhodes and Castel (2008) to conclude that the effect is driven by fluency of processing. However, more recently, Mueller et al. (2014) provided convincing evidence that the font size effect is caused more by an explicit belief about how font size will influence memory, rather than an actual difference in fluency. Specifically, they showed that font size has no measurable behavioral effects on perceptual fluency, and that the font size effect arises in prestudy JOLs, for which processing fluency cannot possibly influence judgments (because the items have not yet been processed). Importantly, all of experiments conducted by Rhodes and Castel (2008) and by Mueller et al. (2014) used a within-subjects manipulation of font size, which therefore directly contrasted the large-font words with the small-font words. Susser, Mulligan, and Besken (2013) demonstrated that a between-subjects manipulation eliminated the font size effect in JOLs (see Yue, Castel, & Bjork, 2013, for a similar finding with word clarity). We would predict that any similar metacognitive illusion will also be eliminated in a study context that fails to directly contrast the two encoding tasks or conditions.

Conclusion

Emotional information is generally remembered differently than neutral information (e.g., Kensinger & Corkin, 2004). Although participants tend to provide higher JOLs for emotional than for neutral information (Hourihan & Bursey, 2016; Nomi, Rhodes, & Cleary, 2013; Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010), these JOLs may be made more on the basis of list-composition factors that make the emotional content distinctive, rather than on the basis of the physiological response associated with encountering an emotional stimulus (e.g., Hamann, 2001). Although emotional events are clearly remembered differently from neutral events (Levine & Edelstein, 2009), it is less clear that lists of emotional words are truly recalled all that differently than lists of neutral words, on the basis of emotional content alone (Talmi, 2013). Research into practical applications, such as eyewitness memory for emotional events, should also examine metamnemonic judgments; predictions of future memory for emotional information may not be as accurate as we might think.

Appendix

Table 6   Word lists used in Experiment 1

	Number	Word	Valence	Arousal	Frequency
Low arousal	843	lazy	4.38	2.65	2.77
	2186	sigh	4.15	2.78	2.24
	339	quiet	5.58	2.82	3.78
	1761	knitting	4.94	2.88	2.06
	416	subdued	4.67	2.90	1.52
	909	nun	4.93	2.93	2.55
	380	seat	4.95	2.95	3.60
	309	pencil	5.22	3.14	2.70
	66	chair	5.08	3.15	3.40
	57	butter	5.33	3.17	3.02
	810	indifferent	4.61	3.18	1.77
	832	kettle	5.22	3.22	2.16
	1218	broom	4.83	3.23	2.39
	825	item	5.26	3.24	2.80
	776	hairpin	5.26	3.27	1.28
	864	mantel	4.93	3.27	1.59
	353	reserved	4.88	3.27	2.41
	658	bland	4.10	3.29	1.75
	685	chin	5.29	3.31	2.81
	651	barrel	5.05	3.36	2.73
High arousal	553	cliff	4.67	6.25	3.04
	2257	squeal	4.67	6.26	2.14
	979	shotgun	4.37	6.27	2.71
	1365	crazy	5.93	6.28	4.14
	1885	nerves	4.28	6.31	2.75
	603	rifle	4.02	6.35	2.87
	2013	pregnant	4.30	6.37	3.42
	904	noisy	5.02	6.38	2.41
	915	obsession	4.52	6.41	2.46
	1273	chase	5.07	6.50	3.22
	1337	conquest	5.85	6.50	1.99
	1391	dare	5.76	6.57	3.45
	1200	boom	5.10	6.67	3.05
	1379	crush	5.90	6.70	2.93
	2463	wolf	5.00	6.70	3.01
	1589	frenzy	4.97	6.86	2.02
	21	anxious	4.81	6.92	2.86
	410	startled	4.50	6.93	2.10
	1708	hysterical	5.29	7.36	2.56
	1524	explosion	5.18	7.93	2.93

"Number" refers to the entry number in the ANEW database (Bradley & Lang, 2010)

Table 7 Word lists used in Experiment 2

	Number	Word	Valence	Arousal	Frequency
Neutral valence	1992	policy	4.50	4.30	3.14
	2091	revert	4.53	4.73	1.56
	416	subdued	4.67	2.90	1.52
	1351	counselor	4.68	4.54	2.71
	1112	assume	4.69	4.97	3.23
	1305	clinic	4.72	4.81	2.85
	1759	kidney	4.86	4.48	2.69
	1498	enzyme	4.90	4.34	1.74
	2331	tendon	4.90	4.90	1.48
	1056	admit	4.93	4.97	3.48
	1893	noise	4.93	4.86	3.25
	2079	repent	4.93	5.11	2.09
	2410	usage	4.93	4.90	1.43
	1372	crocodile	5.00	6.03	2.06
	346	rattle	5.03	4.36	2.24
	1264	cavort	5.10	4.86	0.60
	613	tank	5.16	4.88	3.12
	1796	lizard	5.23	5.13	2.39
	1708	hysterical	5.29	7.36	2.56
	1543	fate	5.43	5.36	3.14
Negative valence	269	maggot	2.06	5.28	1.91
	365	rotten	2.26	4.53	2.95
	124	disgusted	2.45	5.42	1.96
	782	hardship	2.45	4.76	1.89
	1290	cigarette	2.46	5.35	3.13
	1909	odor	2.52	5.13	2.08
	76	coffin	2.56	5.03	2.66
	400	slime	2.61	4.57	1.79
	65	cemetery	2.63	4.82	2.65
	2165	sewage	2.68	5.36	2.15
	272	measles	2.74	5.06	2.03
	885	mosquito	2.80	4.78	1.97
	169	foul	2.81	4.93	2.87
	2007	pout	2.83	5.07	1.75
	1747	jerk	2.86	5.21	3.23
	284	morbid	2.87	5.06	2.12
	704	crime	2.89	5.41	3.56
	1363	cranky	2.90	5.36	2.24
	1087	annoyance	2.97	5.18	1.42
	182	garbage	2.98	5.04	3.12

"Number" refers to the entry number in the ANEW database (Bradley & Lang, 2010).

Table 8 Word pool used in Experiment 3

Number	Word	Valence	Arousal	Frequency
726	discomfort	2.19	4.17	1.91
731	dreary	3.05	2.98	1.97
162	fatigued	3.28	2.64	1.4
779	handicap	3.29	3.81	2.15
1789	limp	3.71	4.29	2.27
277	mildew	3.17	4.08	1.36
367	rusty	3.86	3.77	2.82
368	sad	1.61	4.13	3.51
2357	tired	3.28	2.64	3.76
1164	beg	2.75	5.00	3.42
1400	defeat	2.97	5.63	2.76
1412	dent	2.93	5.69	2.26
1579	foolish	3.54	4.46	2.95
842	lawsuit	3.37	4.93	2.49
1852	miserable	1.55	5.00	3.04
284	morbid	2.87	5.06	2.12
1962	pessimism	3.10	4.74	1.2
365	rotten	2.26	4.53	2.95
18	angry	2.85	7.17	3.48
37	betray	1.68	7.24	2.67
85	controlling	3.80	6.10	2.36
1600	furious	1.96	7.64	2.49
222	hurt	1.90	5.85	4.1
237	jealousy	2.51	6.36	2.47
1906	obsessed	3.60	6.24	2.72
342	rage	2.41	8.17	2.76
970	scalding	2.82	5.95	1.32
57	butter	5.33	3.17	3.02
1761	knitting	4.94	2.88	2.06
1803	lounge	6.60	3.97	2.6
303	paper	5.20	2.50	3.72
2220	snail	4.31	3.86	1.96
2235	sofa	6.53	3.10	2.48
408	square	4.74	3.18	3.21
426	table	5.22	2.92	3.73
2318	tatter	4.53	3.90	N/A
4	activate	5.46	4.86	2.38
1276	cheerleader	6.48	5.67	2.52
1358	coyote	4.86	5.19	2.27
1392	daughter	6.36	5.11	3.94
198	hammer	4.88	4.58	2.8
1759	kidney	4.86	4.48	2.69
2184	shut	4.53	4.87	4.13
2247	spice	6.21	5.62	2.43
1200	boom	5.10	6.67	3.05
1524	explosion	5.18	7.93	2.93
1589	frenzy	4.97	6.86	2.02
539	plane	6.43	6.14	3.69
949	python	4.05	6.18	1.9

Table 8 (continued)

Number	Word	Valence	Arousal	Frequency
2173	shock	4.03	7.45	3.17
2252	sports	6.45	6.23	3.15
410	startled	4.50	6.93	2.1
2371	touch	6.31	6.19	3.88
38	bird	7.27	3.17	3.37
1185	blanket	6.94	3.41	2.82
42	bless	7.19	4.05	3.25
246	kindness	7.82	4.30	2.66
308	peace	7.72	2.95	3.55
315	pillow	7.92	2.97	2.76
320	politeness	7.18	3.74	1.48
333	protected	7.29	4.09	2.74
404	snuggle	7.92	4.16	1.83
629	advantage	6.95	4.76	3.05
105	delight	8.26	5.44	2.46
245	kind	7.59	4.46	4.48
1808	lunch	7.21	5.43	3.73
1831	mastery	6.69	5.62	1.48
888	muffin	6.57	4.76	2.47
304	paradise	8.72	5.12	2.83
1007	taste	6.66	5.22	3.42
468	vacation	8.16	5.64	3.22
503	cash	8.37	7.37	3.57
69	cheer	8.10	6.12	2.98
1513	excite	7.60	7.16	2.04
279	miracle	8.60	7.65	3.13
1917	opportunity	7.41	6.47	3.32
2018	prince	7.03	6.07	3.36
427	talent	7.56	6.27	3.12
438	thrill	8.05	8.02	2.63
2354	tickle	6.86	6.70	2.39

"Number" refers to the entry number in the ANEW database (Bradley & Lang, 2010)

References

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81,* 126–131. doi:10.1037/h0027455

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using "Eigen" and S4 (R package version 1.1-8). Retrieved from http://CRAN.R-project.org/package=lme4

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610–632.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31,* 297–305. doi:10.3758/bf03194388

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Mahwah: Erlbaum.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68. doi:10.1037/0096-3445.127.1.55

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York: Psychology Press.

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

Bradley, M. M., & Lang, P. J. (2010). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings (Technical Report C-2).* Gainesville: University of Florida, Center for Research in Psychophysiology.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasing declarative memory. *Trends in Neurosciences, 21,* 294–299.

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition, 41,* 28–35. doi:10.3758/s13421-012-0249-6

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253. doi:10.1177/014662168300700030

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Categorical or nominal independent variables. In *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed., pp. 302–353). Mahwah, NJ: Erlbaum.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505. doi:10.1080/14640748108400805

Dewhurst, S. A., & Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology, 12,* 541–551. doi:10.1080/095414400750050222

Fraundorf, S. H., Benjamin, A. S., & Watson, D. G. (2013). What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language, 69,* 196–227. doi:10.1016/j.jml.2013.06.003

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language, 63,* 367–386. doi:10.1016/j.jml.2010.06.004

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62,* 1–18. doi:10.1016/j.jml.2009.09.004

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hall, J. F. (1954). Learning as a function of word frequency. *American Journal of Psychology, 67,* 138–140.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5,* 394–400.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 22–34. doi:10.1037/0278-7393.29.1.22

Hourihan, K. L., & Bursey, E. (2016). A misleading feeling of happiness: Metamemory for positive emotional and neutral pictures. *Memory.* doi:10.1080/09658211.2015.1122809

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: The cross-race effect in face memory with social category information. *Memory & Cognition, 41,* 1021–1031. doi:10.3758/s13421-013-0316-7

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review, 1,* 99–113. doi: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,* 1169–1180. doi:10.3758/BF03195800

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

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

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643–656. doi:10.1037/0096-3445.133.4.643

Levine, L. J., & Edelstein, R. S. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition and Emotion, 23,* 833–875. doi:10.1080/02699930902738863

Mandler, G. (1975). *Mind and emotion*. New York: Wiley.

Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 509–527. doi:10.1037/a0014876

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6,* 114–133. doi:10.1177/1745691611400234

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.

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model

perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1287–1306. doi:10.1037/a0036914

Nomi, J. S., Rhodes, M. G., & Cleary, A. M. (2013). Emotional facial expressions differentially influence predictions and performance for face recognition. *Cognition and Emotion, 27,* 141–149. doi:10.1080/02699931.2012.679917

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

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

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

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44,* 695–729. doi:10.1177/0539018405058216

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

Siddiqui, A. P., & Unsworth, N. (2011). Investigating the role of emotion during the search process in free recall. *Memory & Cognition, 39,* 1387–1400. doi:10.3758/s13421-011-0125-9

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,* 1000–1011. doi:10.3758/s13421-013-0323-8

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14,* 455–461.

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science, 22,* 430–436. doi:10.1177/0963721413498893

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

Tauber, S. K., & Dunlosky, J. (2012). Can older adults accurately judge their learning of emotional information? *Psychology and Aging, 27,* 924–933. doi:10.1037/a0028447

Tomaszczyk, J. C., Fernandes, M. A., & MacLeod, C. M. (2008). Personal relevance modulates the positivity bias in recall of emotional pictures in older adults. *Psychonomic Bulletin & Review, 15,* 191–196. doi:10.3758/PBR.15.1.191

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is–and is not–a desirable difficulty: the influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition, 41,* 229–241. doi:10.3758/s13421-012-0255-8

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,* 240–253. doi:10.1016/j.jml.2009.11.004