BRIEF REPORT

The effect of perceptual processing fluency and value on metacognition and remembering

Dillon H. Murphy[1] · Stephen C. Huckins[1] · Matthew G. Rhodes[2] · Alan D. Castel[1]

Abstract

Previous research has indicated that perceptual processing fluency significantly affects metacognitive predictions of performance but not learning outcomes. In the present study, we examined the differential impact of perceptual processing fluency and an item's value on metacognition and recall. We presented participants with words visually and audibly, with each word paired with a point value counting towards participants' scores if recalled. The words were either highly perceptually fluent (large font, loud volume) or less perceptually fluent (small font, low volume). Results revealed that both metacognitive monitoring (JOLs) and recall were sensitive to perceptual processing fluency as well as value, but the magnitude of the effect of value was significantly greater than that of font size. Specifically, high-value words were better remembered than low-value words, regardless of fluency, and participants' judgments mapped onto their selectivity for valuable information. Thus, the current study revealed the differential effects of intrinsic and extrinsic cues on metacognitive monitoring and later remembering such that the cues that can influence monitoring in certain encoding conditions become less impactful when pitted against other intrinsic cues in different encoding conditions.

Keywords Selectivity · Fluency · Metacognition · Monitoring

Introduction

Vision and hearing allow us to perceive and encode information from our surroundings and perceptual fluency is the ease with which these stimuli are processed (Oppenheimer, 2008). For example, large, bolded headlines at the top of newspapers are highly fluent visual stimuli, while loud music or other distinct sounds exemplify highly fluent audio material. In contrast, there are many instances when stimuli are less perceptually fluent, such as when visual information becomes blurry, or when audio becomes faint or muffled. Although learners often expect more fluent information, such as words in large font or loud words to be better remembered, perceived fluency does not always lead to detectable differences in memory (e.g., Ball et al., 2014; Rhodes & Castel, 2008, 2009; see also Kuhlmann et al., 2020).

Metacognition involves the understanding and awareness of one's memory processes (Nelson & Narens, 1990; see also Dunlosky et al., 2016; Nelson, 1996) and is usually considered in terms of monitoring (i.e., awareness of learning) and control (i.e., self-regulation of learning). Metacognitive control processes are typically evaluated via self-regulation of study time and study choices (Egner, 2017; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999), and these processes are informed by metacognitive monitoring. First introduced by Arbuckle and Cuddy (1969), metacognitive monitoring is often evaluated via judgments of the likelihood of later remembering information, also known as a "judgment of learning" (JOLs; see Rhodes, 2016, for a review). When JOLs are sensitive to cues that affect memory performance and unaffected by those that have minimal effects, metacognitive monitoring assessments are generally accurate (e.g., Murphy et al., 2021). However, when JOLs do not match learning, this dissociation can provide important theoretical and practical insights (see Yang et al., 2021).

Monitoring judgments usually occur after an item is studied and are often informed by the cues available during encoding. To better understand the types of cues that influence JOLs, Koriat (1997) proposed a cue-utilization

✉ Dillon H. Murphy
   dmurphy8@ucla.edu

1 Department of Psychology, University of California, Los Angeles, CA 90095, USA

2 Department of Psychology, Colorado State University, Fort Collins, CO, USA

🖄 Springer

framework whereby three types of cues inform monitoring judgments: *intrinsic cues,* characteristics of to-be-remembered information that affect or are believed to affect learning such as word relatedness; *extrinsic cues*, the encoding operations employed by the learner as well as factors that pertain either the circumstances of encoding or testing such as study time or the type of test (i.e., recall versus recognition); and *mnemonic cues*, one's prior experience with stimuli, such as the ease with which an item comes to mind in response to a cue. JOL accuracy is a function of the degree to which such judgments are sensitive to cues that affect later remembering (Dunlosky & Matvey, 2001; Tiede & Leboe, 2009).

In addition to a variety of cues and heuristics that likely influence predictions of future recall probability, JOLs may also be theory-based (information-based; analytic-processing theory) and mnemonic-based (experience-based). According to analytic-processing theory, theory- or belief-based judgments are informed by analytic inferences about cues that may be predictive of later remembering (e.g., a highly salient item may be believed to be easier to process and thus expected to be better remembered; Dunlosky et al., 2015; Mueller et al., 2016; see also Blake & Castel, 2018). In contrast, mnemonic-based (experience-based) judgments are informed by implicit inferences of memorability based on subjective experience when studying a to-be-remembered item (e.g., encoding fluency; Koriat & Bjork, 2006).

While JOLs are generally accurate, such that participants predict their performance better than chance (Rhodes, 2016), there are some instances when judgments show little correspondence with later remembering. For example, Rhodes and Castel (2008) demonstrated that participants judged words in large font as more likely to be remembered than words in a small font, but font size did not affect participants' actual memory performance (see also Besken & Mulligan, 2013; Kornell et al., 2011; Mueller & Dunlosky, 2016; Yue et al., 2013; but see Sungkhasettee et al., 2011). Similar illusory effects of perceptual processing fluency on participants' JOLs have also been found for auditory information. Specifically, Rhodes and Castel (2009) found that loud words were rated as more likely to be remembered than quieter words, but again, perceptual processing fluency did not affect later remembering (see also Foster & Sahakyan, 2012; Frank & Kuhlmann, 2017; Soderstrom & Rhodes, 2014). Thus, in some instances, people may falsely believe that more perceptually fluent information is more likely to be recalled, while less perceptually fluent information is less likely to be remembered.

Although perceptual processing fluency may not impact later remembering in some instances (e.g., Ball et al., 2014; Rhodes & Castel, 2008, 2009), when to-be-remembered words are paired with point values that count towards participants' scores if recalled, participants generally recall valuable information better than low-value information (Ariel et al., 2015; Castel et al., 2002, 2007, 2013; Nguyen et al., 2019; see Madan, 2017, for review). Thus, absent variations in perceptual processing fluency, participants prioritize value when guiding the encoding and retrieval process. However, when to-be-remembered information varies in both value and fluency, Soderstrom and McCabe (2011) demonstrated that item value, as well as processing fluency (as manipulated by word pair relatedness), can influence metacognitive monitoring (see also Koriat et al., 2006). Specifically, participants can engage in agenda-based monitoring (similar to the agenda-based regulation model of study time; see Ariel et al., 2009) by using both value and processing fluency as cues to inform JOLs and subsequent study time, indicating that multiple cues can be used simultaneously to inform JOLs.

When faced with multiple cues at encoding, if participants show a weak relationship between what they expect to remember and later recall, this metacognitive disconnect could result in the forgetting of valuable information. Rather than incorporating a single cue such as processing fluency or value in their metacognitive monitoring judgments, participants should engage in *responsible remembering* (see Murphy & Castel, 2020, 2021a, 2021b) by simultaneously incorporating multiple cues but also differentially *weighting* these cues (i.e., *cue-weighting*, see Bröder & Undorf, 2019; Koriat, 1997; Undorf & Bröder, 2020; Undorf et al., 2018). Specifically, responsible remembering mechanisms may allow for the strategic encoding of important information to maximize memory utility and avoid forgetting valuable information, despite variation in encoding or perceptual processing fluency. Thus, when multiple cues are present, participants should differentially prioritize the encoding and recall of an item based on its value (an extrinsic cue) rather than its processing fluency (an intrinsic cue[1]).

The current study

Previous work indicates that more perceptually fluent materials are judged as more likely to be remembered but may not be better recalled (e.g., Ball et al., 2014; Rhodes & Castel, 2008, 2009). However, other work indicates that highly related word pairs (high processing fluency) as well as valuable items are given higher JOLs and are more likely to be remembered compared to less fluent or low-value information (e.g., Koriat et al., 2006; Soderstrom & McCabe, 2011). In the current study, we examined how perceptual processing

---

[1] Perceptual processing fluency may lead to differences in the interpretation of ease of processing (i.e., an explicit theory regarding how fluency affects recall), which could also serve as a *mnemonic cue*.

fluency and the value of to-be-remembered information differentially affect monitoring of learning and subsequent remembering. We presented participants with words either visually or audibly, with some words being highly perceptually fluent (large font or loud) and others being less fluent (smaller font or quieter). Additionally, each word was paired with either a high or a low point value counting towards participants' scores if later recalled.

Accordingly, the present experiments allowed us to examine the influence of both value and perceptual processing fluency on metacognitive monitoring and later remembering, providing insight regarding the notion of *cue-weighting*, whereby multiple cues are considered when forming JOLs (Bröder & Undorf, 2019; Koriat, 1997; Undorf et al., 2018; Undorf & Bröder, 2020), as well as theoretical frameworks suggesting that fluency can guide JOLs, despite value influencing recall. Specifically, participants may override perceptual processing fluency as a cue and use value as a stronger indicator of future recall. Thus, we expected participants to be selective for high-value words despite some being less perceptually fluent (small font or quiet). Such selectivity might map onto participants' metacognitive judgments such that JOLs more strongly correspond to value than perceptual processing fluency, suggesting that value may be prioritized when making metacognitive judgments in the face of perceptual processing fluency.

Experiment 1a

In Experiment 1a, we examined how value and font size differentially impact metacognitive judgments and recall. We presented participants with eight words paired with high (5) point values and eight words paired with low (1) point values. Half of the high- and low-value words were presented in a large font (more perceptually fluent), and half were presented in small font (less perceptually fluent). After the presentation of a word, participants estimated the likelihood of later remembering it (JOL). We expected participants to be selective for valuable items, regardless of fluency, and for JOLs to map on to their selectivity, consistent with engaging in responsible remembering. Specifically, the importance of some to-be-remembered information may override other cues that, in the absence of value, can guide the encoding and retrieval process.

Method

Participants After exclusions, participants were 51 undergraduate students (M_{age} = 19.10 years, SD_{age} = 1.37) 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 analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (participants were told they would still receive credit if they cheated). This exclusion process resulted in one exclusion. An a priori power analysis indicated that for a 2 (font size: small, large) × 2 (value: 1, 5) repeated-measures, within-subjects ANOVA, with a low correlation between repeated-measures, assuming alpha = .05, power = .80, 38 participants would be needed to reliably detect a medium effect size (η^2 = .10).

Materials and procedure Participants were presented with a series of to-be-remembered words with each word paired with a point value indicating how much the word was "worth." Point values were either 1 (low) or 5 (high), and the order of the point values within lists was randomized. Half of the words were presented in a large (48 point) font, and half were presented in a small (12-point) font such that each list included four large valuable words, four small valuable words, four large low-value words, and four small low-value words. Regardless of the size of each to-be-remembered word, the point values paired with each word were presented in a neutral size (30-point) font. Participants were told that their score would be the sum of the associated values of the words they recalled and that they should try to maximize their score but were not given any information regarding word size.

Word-value pairs (e.g., ability : 5) were separated by a colon, presented simultaneously for 4 s each, and were taken from Rhodes and Castel (2009). The to-be-remembered words were between four and eight letters (*M* = 6.19, *SD* = 1.27) and on the log-transformed Hyperspace Analogue to Language frequency scale (with lower values indicating lower frequency in the English language and higher values indicating higher frequency), ranged from 7.22 to 13.67 and averaged a score of 10.15 (*SD* = .88). In terms of concreteness (with lower values indicating lower concreteness and higher values indicating higher concreteness), words ranged from 1.19 to 5.00 and averaged a score of 3.31 (*SD* = 1.00). Words were classified according to the English Lexicon Project website (Balota et al., 2007).

After each word was presented, participants were asked to estimate the likelihood of correctly recalling it on a later test (JOL). Participants answered with a number between 0 and 100, with 0 meaning they definitely would not remember the word and 100 meaning they definitely would remember the word. Participants were given as much time as they needed to make their judgments. After the presentation of all 16 words, participants were given a 1-min free-recall test in which they had to recall as many words as they could from the current list (they did not need to recall point values). Immediately following the recall period, participants were informed of their score for that list but were not given
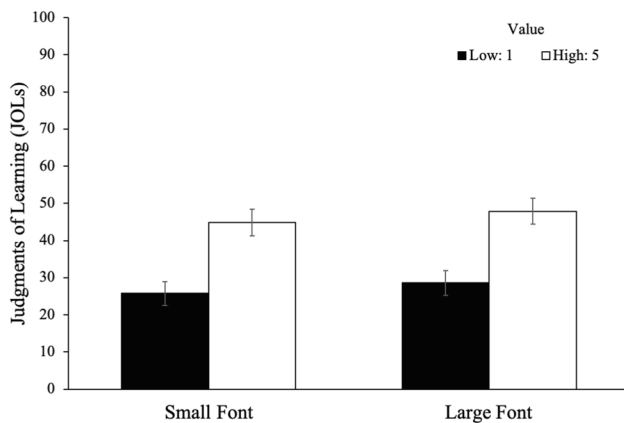
Fig. 1 Judgments of learning (JOLs) as a function of font size and value in Experiment 1a. Error bars reflect the standard error of the mean

Fig. 2 Recall as a function of font size and value in Experiment 1a. Error bars reflect the standard error of the mean

feedback about specific items. This process was repeated for a total of eight study-test cycles.

Results

Judgments of learning (JOLs) as a function of font size and value are shown in Fig. 1. To investigate differences in JOLs as a function of font size and value, a 2 (size: small, large) × 2 (value: 1, 5) repeated-measures ANOVA revealed that large words ($M = 38.24$, $SD = 21.89$) were rated as more likely to be remembered than small words ($M = 35.32$, $SD = 21.34$), [$F(1, 50) = 10.44$, $p = .002$, $\eta^2 = .17$]. Additionally, high-value words ($M = 46.34$, $SD = 24.89$) were rated as more likely to be remembered than low-value words ($M = 27.22$, $SD = 23.14$), [$F(1, 50) = 38.61$, $p < .001$, $\eta^2 = .44$]. However, font size did not interact with value [$F(1, 50) = .04$, $p = .841$, $\eta^2 < .01$].

Recall as a function of font size and value is shown in Fig. 2. A 2 (size: small, large) × 2 (value: 1, 5) repeated-measures ANOVA showed that large words ($M = .52$, $SD = .17$) were recalled better than small words ($M = .48$, $SD = .17$), [$F(1, 50) = 4.68$, $p = .035$, $\eta^2 = .09$]. Additionally, high-value words ($M = .63$, $SD = .19$) were recalled better than low-value words ($M = .36$, $SD = .22$), [$F(1, 50) = 59.64$, $p < .001$, $\eta^2 = .54$], but size did not interact with value [$F(1, 50) = .16$, $p = .693$, $\eta^2 < .01$].

Because participants' JOLs were assessed as a probability, or percentage likelihood (same scale as the probability of recall), we calculated measures of *absolute* and *relative accuracy* (see Higham et al., 2016; Rhodes, 2016). Absolute accuracy (i.e., *calibration*), captures the overall relationship between judgments and performance and is calculated as the difference between mean judgments and the percentage of items recalled. A scor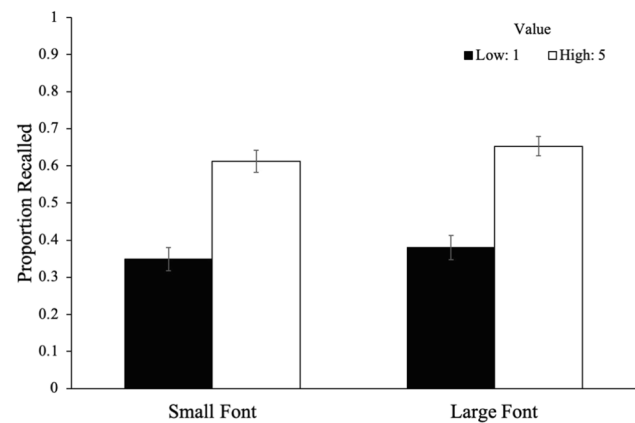e of zero would exemplify perfect calibration (i.e., a direct correspondence between participants' predictions and actual recall). A 2 (size: small, large) × 2 (value: 1, 5) repeated-measures ANOVA on calibration scores showed that participants were similarly calibrated for large words ($M = -13.41$, $SD = 17.86$) as for small words ($M = -12.72$, $SD = 17.67$), [$F(1, 50) = .29$, $p = .593$, $\eta^2 = .01$]. However, participants were more overconfident in their memory for high-value words ($M = -16.90$, $SD = 21.33$) than low-value words ($M = -9.23$, $SD = 17.33$), [$F(1, 50) = 8.96$, $p = .004$, $\eta^2 = .15$]. Size did not interact with value [$F(1, 50) = .08$, $p = .781$, $\eta^2 < .01$].

Relative accuracy (i.e., *resolution*) measures participants' ability to discriminate between items that will be or will not be remembered and is often measured via Gamma correlations between a given item's JOL and whether the given item was recalled (see Masson & Rotello, 2009 for alternative approaches). A strong positive correlation between judgments and performance would indicate that an individual remembers what they say they will remember while a strong negative correlation would indicate that an individual forgets what they say they will remember. We computed Gamma correlations for each participant and examined resolution as a function of font size and value using a 2 (size: small, large) × 2 (value: 1, 5) repeated-measures ANOVA. Overall, participants were similarly relatively accurate for large words ($M = .37$, $SD = .43$) and small words ($M = .34$, $SD = .44$), [$F(1, 48) = .16$, $p = .691$, $\eta^2 < .01$]. Additionally, participants were similarly relatively accurate for high-value words ($M = .32$, $SD = .45$) as low-value words ($M = .39$, $SD = .43$), [$F(1, 48) = 2.56$, $p = .117$, $\eta^2 = .05$]. Size did not interact with value [$F(1, 48) = .91$, $p = .346$, $\eta^2 = .02$].

Discussion

In Experiment 1a, we investigated how perceptual processing fluency and value differentially affected metacognitive

monitoring and memory performance. Results revealed that participants' judgments and recall were sensitive to both font size and word value (but there were no interactions between the two). Thus, participants engaged in responsible remembering by recalling valuable words, regardless of font size, and were generally metacognitively aware of this pattern.

Experiment 1b

To support the results of Experiment 1a, we report a conceptual replication of Experiment 1a.[2] Participants were presented with words differing in font size and value; however, rather than a value of 5 serving as the high-value word, the high-value words were worth 10 points. Additionally, rather than studying a single list followed by an immediate recall test (and several study-test cycles), participants studied a single list of words before completing a delayed recall test. Again, the manipulation of value (either 1 or 10 points) was crossed with font size (small or large font) such that large and small words were presented equally often with high- and low-value words.

Method

Participants Participants were 32 Colorado State University psychology students (M_{age} = 18.44 years, SD_{age} = 1.22). Participants were tested individually or in groups of up to eight participants and received course credit for their participation. A sensitivity analysis indicated that for a 2 (font size: small, large) × 2 (value: 1, 10) within-subjects ANOVA, assuming alpha = .05, power = .80, and a low correlation (r = .10) between repeated-measures, the smallest effect size the design could reliably detect is η^2 = .11.

Materials The to-be-remembered words were between four and eight letters (M = 6.00, SD = 1.27) and on the log-transformed Hyperspace Analogue to Language frequency scale, ranged from 7.22 to 11.25, and averaged a score of 9.57 (SD = .86). In terms of concreteness, words ranged from 1.52 to 4.72 and averaged a score of 3.22 (SD = .91). Half of the words were assigned a value of "10" and half were assigned a value of "1." This manipulation of value was crossed with font size, such that half of the words in large or small font were presented as high-value (10 point) items and half were presented as low-value (1 point) items.

---

[2] After creating and submitting the manuscript, the experimenters reached out to the third author who noted that he had conducted a highly similar, unpublished experiment several years ago. These data are now incorporated as Experiment 1b.

Procedure The procedure was similar to Experiment 1a. However, following each item, participants were only given 4 s to record their JOL on a sheet of paper provided. Additionally, words were presented for 5 s each in a fixed random order (with the condition that no more than three items of the same font size were presented consecutively). Immediately following the study list participants engaged in a filler task for approximately 5 min that required them to write down the states of the USA. Finally, participants were instructed to recall as many of the words as they could remember on a blank sheet of paper provided. Participants were given 4 min for the free-recall task.

Results

JOLs as a function of font size and value are shown in Fig. 3. To investigate differences in JOLs as a function of font size and value, a 2 (size: small, large) × 2 (value: 1, 10) repeated-measures ANOVA revealed that, overall, large words (M = 44.50, SD = 16.79) were rated as more likely to be remembered than small words (M = 39.63, SD = 16.40), [$F(1, 31)$ = 8.48, p = .007, η^2 = .22]. Additionally, high-value words (M = 52.32, SD = 18.74) were rated as more likely to be remembered than low-value words (M = 31.71, SD = 18.55), [$F(1, 31)$ = 35.28, p < .001, η^2 = .53]. However, font size did not interact with value [$F(1, 31)$ = 1.61, p = .214, η^2 = .05].

Recall as a function of font size and value is shown in Fig. 4. A 2 (size: small, large) × 2 (value: 1, 10) repeated-measures ANOVA showed that large words (M = .17, SD = .07) were recalled better than small words (M = .13, SD = .09), [$F(1, 31)$ = 6.84, p = .014, η^2 = .18]. Additionally, high-value words (M = .21, SD = .12) were recalled better than low-value words (M = .09, SD = .06), [$F(1, 31)$ =
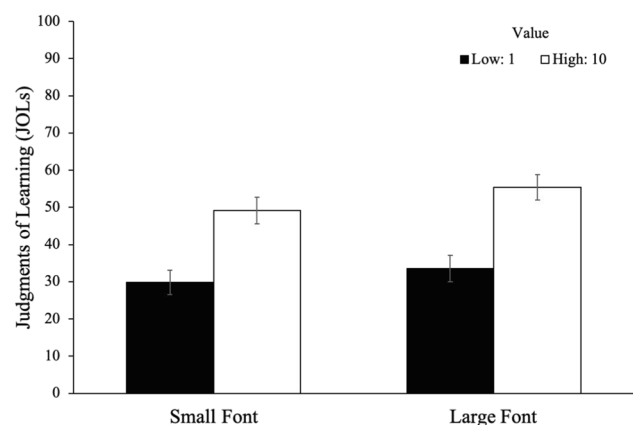

Fig. 3 Judgments of learning (JOLs) as a function of font size and value in Experiment 1b. Error bars reflect the standard error of the mean
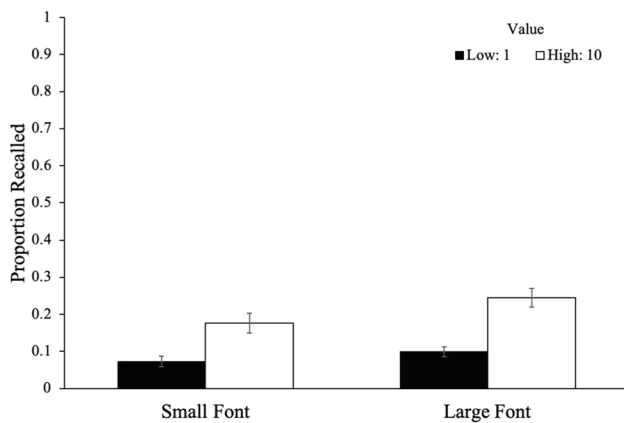
Fig. 4 Recall as a function of font size and value in Experiment 1b. Error bars reflect the standard error of the mean

27.22, $p < .001$, $\eta^2 = .47$]. Size did not interact with value [$F(1, 31) = 1.82$, $p = .187$, $\eta^2 = .06$].

To examine differences in calibration as a function of font size and value, a 2 (size: small, large) × 2 (value: 1, 10) repeated-measures ANOVA showed that participants were similarly calibrated for large words ($M = 27.32$, $SD = 20.88$) and small words ($M = 27.13$, $SD = 19.76$), [$F(1, 31) = .01$, $p = .924$, $\eta^2 < .01$]. However, participants were more underconfident in their memory for high-value words ($M = 31.23$, $SD = 24.75$) than low-value words ($M = 23.12$, $SD = 18.27$), [$F(1, 31) = 5.13$, $p = .031$, $\eta^2 = .14$]. Size did not interact with value [$F(1, 31) = .24$, $p = .628$, $\eta^2 = .01$].

To examine differences in resolution as a function of font size and value, a 2 (size: small, large) × 2 (value: 1, 10) repeated-measures ANOVA showed that participants were similarly relatively accurate for large words ($M = .38$, $SD = .28$) as for small words ($M = .45$, $SD = .37$), [$F(1, 11) = 1.94$, $p = .192$, $\eta^2 = .15$]. Additionally, participants were similarly relatively accurate for high-value words ($M = .32$, $SD = .50$) as for low-value words ($M = .32$, $SD = .56$), [$F(1, 11) = .39$, $p = .547$, $\eta^2 = .03$]. Size did not interact with value [$F(1, 11) = 1.46$, $p = .252$, $\eta^2 = .12$].

Discussion

The data from Experiment 1b largely replicated Experiment 1a. Specifically, both size and value significantly influenced JOLs such that larger words and high-value words were given higher JOLs than small words and low-value words, respectively, and no interaction was present. This is borne out by the recall data, as large words and high-value words were also more likely to be recalled than small words and low-value words, respectively.

Experiment 2

In Experiment 1, participants were more sensitive to value than font size, and their JOLs were generally consistent with recall. In Experiment 2, we further examined how value and perceptual processing fluency differentially impact recall and metacognitive judgments. Rather than presenting words visually, the to-be-remembered words were presented audibly (with the corresponding values still appearing on the screen). Specifically, low- and high-value words were presented audibly at either a low volume (less fluent) or a louder volume (fluent) in a 2 × 2 design. We expected to observe a similar pattern of results as seen in Experiment 1 such that participants selectively remember valuable items at the expense of low-value items, but also better remember highly perceptually fluent words compared with less fluent words, and for these trends to map on to participants' metacognitive judgments.

Methods

Participants Participants were 51 undergraduate students ($M_{age} = 19.33$ years, $SD_{age} = 1.96$) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (participants were told that they would still receive credit if they cheated). This exclusion process resulted in zero exclusions. A sensitivity analysis indicated that for a 2 (volume: low, high) × 2 (value: 1, 5) repeated-measures ANOVA, with a low correlation between repeated-measures, assuming alpha = .05, power = .80, the smallest effect size the design could reliably detect is $\eta^2 = .07$.

Materials and procedure The materials and procedure were similar to Experiment 1a; however, words were presented audibly, and values appeared simultaneously on the screen in size 30 font. The lower-volume words were presented at a conversational volume, while the high-volume words were presented at a louder volume (about three times louder; see Rhodes & Castel, 2009). Participants were not required to use headphones (i.e., they could use their computer's speaker) but were required to complete an audio test before beginning the task.

Results

JOLs as a function of word volume and value are shown in Fig. 5. To investigate differences in JOLs as a function of volume and value, a 2 (volume: quiet, loud) × 2 (value: 1, 5)
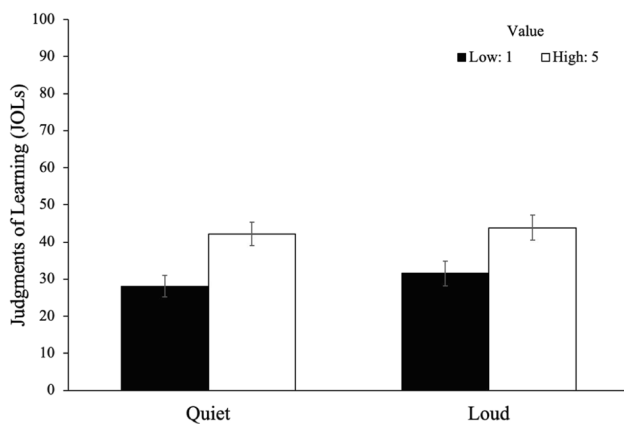
Fig. 5 Judgments of learning (JOLs) as a function of word volume and value in Experiment 2. Error bars reflect the standard error of the mean

repeated-measures ANOVA revealed that loud words ($M = 37.73$, $SD = 21.42$) were rated as more likely to be remembered than quiet words ($M = 35.13$, $SD = 18.79$), $[F(1, 50) = 10.40$, $p = .002$, $\eta^2 = .17]$. Additionally, high-value words ($M = 43.00$, $SD = 22.89$) were rated as more likely to be remembered than low-value words ($M = 29.86$, $SD = 22.00$), $[F(1, 50) = 20.76$, $p < .001$, $\eta^2 = .29]$. Moreover, fluency interacted with value $[F(1, 50) = 5.37$, $p = .025$, $\eta^2 = .10]$ such that JOLs for loud, low-value words were greater than quiet, low-value words $[t(50) = 4.06$, $p < .001$, $d = .57]$.

Recall as a function of word volume and value is shown in Fig. 6. A 2 (volume: quiet, loud) × 2 (value: 1, 5) repeated-measures ANOVA showed that loud words ($M = .47$, $SD = .16$) were recalled better than quiet words ($M = .43$, $SD = .16$), $[F(1, 50) = 10.32$, $p = .002$, $\eta^2 = .17]$. Additionally, high-value words ($M = .53$, $SD = .18$) were recalled better than low-value words ($M = .38$, $SD = .20$), $[F(1, 50)$
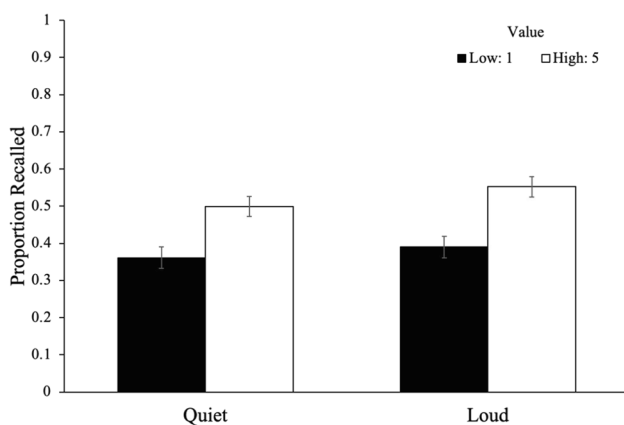

Fig. 6 Recall as a function of word volume and value in Experiment 2. Error bars reflect the standard error of the mean

$= 21.30$, $p < .001$, $\eta^2 = .30]$. Volume did not interact with value $[F(1, 50) = 1.30$, $p = .259$, $\eta^2 = .03]$.

To examine differences in calibration as a function of word volume and value, a 2 (volume: quiet, loud) × 2 (value: 1, 5) repeated-measures ANOVA showed that participants were similarly calibrated for loud words ($M = -9.39$, $SD = 18.99$) as for quiet words ($M = -7.91$, $SD = 19.02$), $[F(1, 50) = 1.95$, $p = .169$, $\eta^2 = .04]$. Additionally, participants were similarly calibrated for high-value words ($M = -9.57$, $SD = 21.27$) and for low-value words ($M = -7.73$, $SD = 18.90$), $[F(1, 50) = .75$, $p = .390$, $\eta^2 = .02]$. Volume did not interact with value $[F(1, 50) = 3.48$, $p = .068$, $\eta^2 = .07]$.

To examine differences in resolution as a function of word volume and value, a 2 (size: small, large) × 2 (value: 1, 5) repeated-measures ANOVA showed that participants were similarly relatively accurate for loud words ($M = .21$, $SD = .45$) as for quiet words ($M = .28$, $SD = .38$), $[F(1, 45) = 2.55$, $p = .117$, $\eta^2 = .05]$. Additionally, participants were similarly relatively accurate for high-value words ($M = .22$, $SD = .39$) as for low-value words ($M = .26$, $SD = .44$), $[F(1, 45) = .39$, $p = .537$, $\eta^2 = .01]$. Volume did not interact with value $[F(1, 45) = .04$, $p = .845$, $\eta^2 < .01]$.

Discussion

In Experiment 2, we further explored the effect of value and perceptual processing fluency on metacognitive judgments and recall using audible stimuli rather than visually presented words. Results generally replicated Experiment 1 such that higher valued and louder words were recalled better than low-value and quieter words, and JOLs were consistent with this pattern (but there were no interactions between fluency and value). Thus, participants engaged in responsible remembering by prioritizing valuable information, even when less perceptually fluent, and JOLs were generally sensitive to cues that were diagnostic of later recall.

Meta-analysis of experiments

To fully illustrate the complete set of experiments, we report a small-scale, fixed-effects meta-analysis based on all of the data collected. For each experiment, we calculated a mean effect size (Cohen's d[3]) for one factor, while collapsing across the other factor (e.g., the meta-analysis of value collapsed across fluency), and also report 95% confidence intervals for each effect size. Cumming and Finch (2005) note that confidence intervals for effect sizes only need to have 50% nonoverlap to

---

[3] This meta-analysis uses Cohen's d as the metric of effect size and while the body of the paper uses eta squared. Since these are not the same metrics (although they can be converted), they generate different confidence intervals.
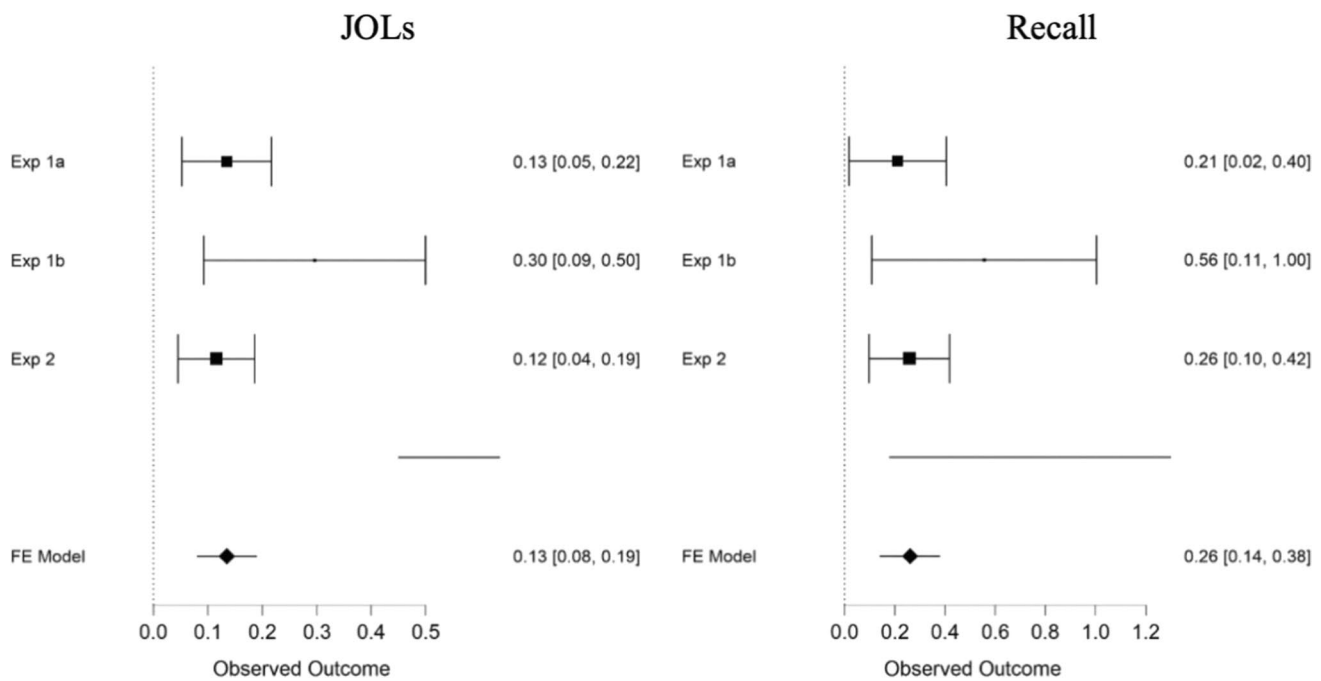
Fig. 7 Forest plots of judgments of learning (JOLs) (**left**) and recall (**right**) for high- versus low-fluency items with 95% confidence intervals. *Note*: FE Model = mean weighted effect size for a fixed effects model; Observed outcome = Cohen's *d*

reach significance with an alphas level of .05. Given the use of repeated-measures designs, we also accounted for the correlation between the two measures, using Cohen's drm (Lakens, 2013). Aggregate effect sizes reported are weighted by sample size (cf. Hedges & Olkin, 1985), and all analyses were conducted using Comprehensive Meta-Analysis Version 2.0 (Borenstein et al., 2005).[4] As can be seen in the forest plot in Fig. 7, fluency yielded reliable effects on both JOLs (left panel; *d* = .13) and recall (right panel; *d* = .26). However, as shown in Fig. 8, value resulted in far stronger effects on JOLs (left panel; *d* = .75) and recall (right panel; *d* = 1.05).[5]

General discussion

Perceptual cues are frequent indicators that people should attend to particular stimuli. For example, in newspaper headlines, people are often drawn to the more fluent, large font

headlines, and people might expect to better remember this information. However, there is some ambiguity surrounding the conditions under which perceptual processing fluency influences subsequent remembering. Some prior work has not detected significant effects of perceptual processing fluency on later remembering using word lists (e.g., Rhodes & Castel, 2008, 2009) as well as more naturalistic stimuli (e.g., Ball et al., 2014) while other research has revealed a small effect of fluency on remembering (see Halamish, 2018; Luna et al., 2018; Price et al., 2016). Specifically, although no single study has reported a memory advantage for large font relative to small font items, Luna et al.'s (2018) meta-analysis of 28 effect sizes showed that a very small memory advantage (*d* unbiased = .08) appears to emerge in favor of large font sizes.

In the absence of variation in perceptual processing fluency, previous work has indicated that people use value to guide encoding and retrieval processes (Ariel et al., 2015; Castel et al., 2002, 2007, 2013; Nguyen et al., 2019; see Madan, 2017, for review). Additionally, previous work has indicated that both word pair relatedness (a form of processing fluency) and value lead to increased JOLs and recall (Koriat et al., 2006; Soderstrom & McCabe, 2011). For example, Rhodes and Castel (2008) presented participants with related and unrelated words to illustrate that font size is given much less weight if other, more compelling cues are available. However, it was previously unclear if a word's value exerted similar effects on recall.

---

[4] We note that the present meta-analysis may have some limitations (such as issues regarding the dependence of variables) and future research should further examine this issue.

[5] Meta-analyses assume each effect size is sampled from independent groups, making a direct comparison of mean effect sizes for value versus fluency statistically untenable given that the effect sizes are derived from the same individuals. Nonetheless, an exploratory analysis comparing the magnitude of effect sizes for value versus fluency confirms that value exerted far stronger effects than fluency on JOLs ($Q_{Between}$ = 40.618, $p < .001$) and recall ($Q_{Between}$ = 27.920, $p < .001$).
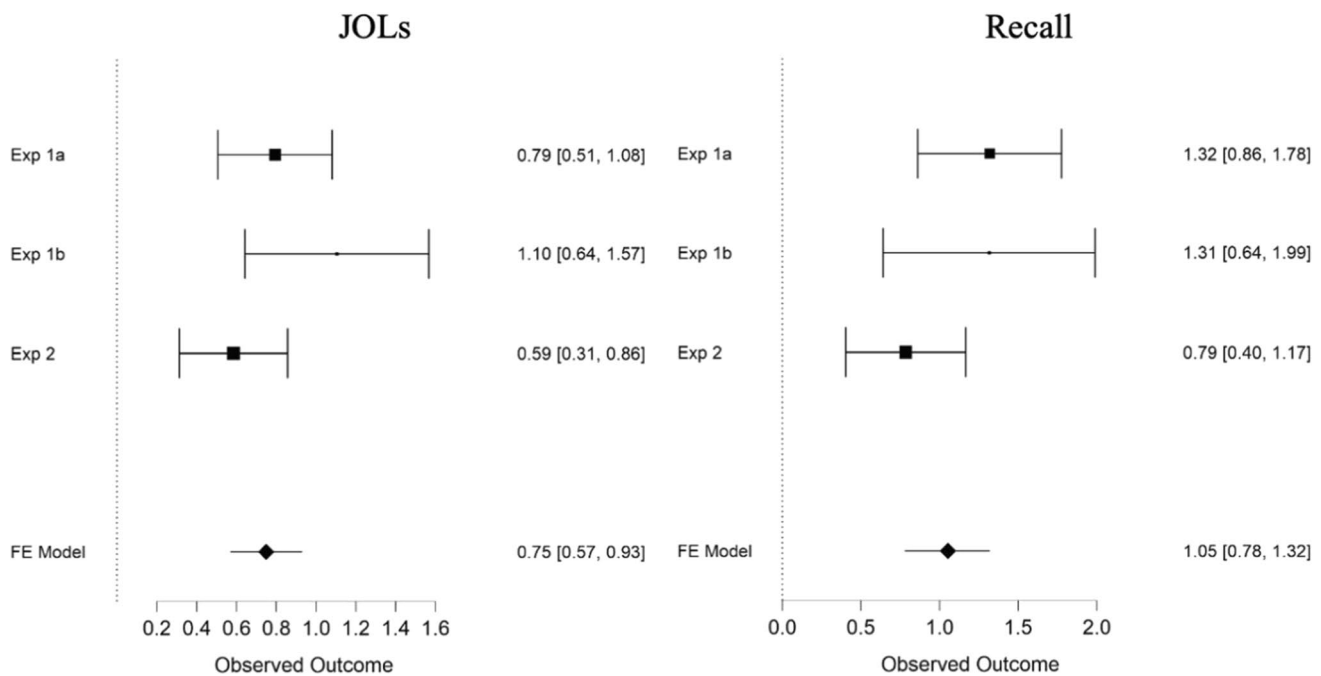
Fig. 8 Forest plots of judgments of learning (JOLs) (**left**) and recall (**right**) for high- versus low-value items with 95% confidence intervals. *Note*: FE Model = mean weighted effect size for a fixed effects model; Observed outcome = Cohen's *d*

In the current study, we presented participants with high- and low-value words of varying levels of perceptual processing fluency (i.e., large font, small font; loud volume, quiet volume) to determine how value and perceptual processing fluency differentially impact metacognitive monitoring and later remembering. Results revealed that increased perceptual processing fluency led to enhanced recall and participants also selectively remembered valuable information at the expense of low-value information, and JOLs mapped onto participants' selectivity (but generally, there were no significant differences in measures of metacognitive accuracy). Thus, the present study is consistent with the idea of cue-weighting (Bröder & Undorf, 2019; Koriat, 1997; Undorf et al., 2018; Undorf & Bröder, 2020) whereby multiple factors inform metacognitive judgments. However, we extend the findings of Soderstrom and McCabe (2011) by indicating that although multiple cues can be used simultaneously to inform JOLs and later remembering, the magnitude of the effect of intrinsic and extrinsic cues on recall can differ and are likely informed by participants' goals. Specifically, as in the current study, fluency may be superseded by a more diagnostic cue like value if the cue is task-relevant, but fluency can still be a sufficiently powerful cue that continues to influence JOLs regardless of the value of the word.

In Koriat's (1997) cue-utilization framework, perceptual processing fluency exemplifies an *intrinsic cue*: characteristics of to-be-remembered information that affect or

are believed to affect memory performance. In terms of an item's value or importance, this *extrinsic cue* serves as a factor that influences the encoding operations employed by participants to maximize memory utility. In the present study, the influence of an item's perceptual processing fluency on JOLs likely reflects a form of analytical processing. Specifically, beliefs or theories of how font size and word volume impact later remembering were used as a diagnostic cue of memorability (i.e., belief-based JOLs). However, when competing with an extrinsic cue such as the item's value, participants more heavily weighed the value or importance of remembering an item than its fluency, and recall mapped onto this trend.

Participants' prioritization of high-value words compared to low-value words, as well as their metacognitive awareness of such selectivity, is consistent with the notion of *responsible remembering*: how our memory allows for the strategic allocation of attention toward important information to avoid consequences for forgetting (Murphy & Castel, 2020, 2021a, 2021b). The present results provide some novel insight regarding how participants may (accurately) prioritize point value (an extrinsic cue) in the presence of competing cues such as perceptual processing fluency. Thus, it may be of strategic benefit to prioritize items in memory according to their value, rather than their ease of processing, and participants may be metacognitively aware of these processes during encoding.

The pattern of JOLs in the present experiments revealed that both intrinsic (perceptual processing fluency) and extrinsic cues (value or importance) can simultaneously inform JOLs but not to the same extent. Although we demonstrated that the magnitude of the effect of value on recall was greater than that of font size, there could be circumstances when an item's font size is highly related to its value. For example, in textbooks, each section usually has a heading in large font (indicating the overarching important theme of the section), important concepts are often referred to multiple times, and important vocabulary words are often bolded. Thus, in more applied settings, value and perceptual processing fluency may not necessarily need to be pitted against each other but rather be highly correlated, and in these instances, participants should engage in strategic metacognitive mechanisms to prioritize this fluent, important information (see Ball et al., 2014; Diemand-Yauman et al., 2011).

There are also instances where important information may be less fluent. For example, warning labels on children's toys and medications' potential side effects are often listed in a small, less perceptually fluent font. However, the present study suggests that this valuable information can still be well remembered, despite being less fluent (see also Hargis & Castel, 2018). Again, this exemplifies responsible remembering such that people can overcome memory obstacles and engage in strategic metacognitive mechanisms to better remember information that might be important later or with potential consequences if forgotten (see Murphy & Castel, 2021a, 2021b).

With the current abundance of auditory and visual learning technologies, understanding how the perceptual processing fluency of information of various importance impacts learning and metacognition is crucial (see Reber & Greifeneder, 2017). The current study indicates that while educators should provide students with more fluent methods of presenting course material to enhance learning outcomes, it may be more important to provide detailed study guides illustrating the relative value or importance of various course concepts to optimize memory for valuable information. The present study also extends Koriat's (1997) cue utilization framework by revealing the differential effects of intrinsic and extrinsic cues on metacognitive monitoring and later remembering. Specifically, not all cues are created equal; some cues that can influence metacognitive monitoring in certain encoding conditions become less impactful when pitted against other cues in different encoding conditions, particularly when cues are relevant to a learner's goals.

Declarations

References

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126-131.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology. General, 138*, 432-447.

Ariel, R., Price, J., & Hertzog, C. (2015). Age-related associative memory deficits in value-based remembering: The contribution of agenda-based regulation and strategy use. *Psychology and Aging, 30*, 795-808.

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Perceptual processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied, 20*, 336-348.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445-459.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition, 41*, 897-903.

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgements of learning: Assessing and altering the direct effects of belief. *Acta Psychologica, 186*, 27-38.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis 2.0. Biostat.

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta psychologica, 197*, 153-165.

Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition, 30*, 1078-1085.

Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition, 35*, 689-700.

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging, 28*, 232-242.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170-180.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118*, 111-115.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1180-1191.

Dunlosky, J., Mueller, M., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning.

In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger, III (Eds.), Remembering: Attributions, processes, and control in human memory (pp. 46-63). Psychology Press.

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* (p. 23-37). Oxford University Press.

Egner, T. (2017). *The Wiley handbook of cognitive control*. Wiley Blackwell.

Foster, N. L., & Sahakyan, L. (2012). Metacognition influences item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1309-1324.

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 680-693.

Halamish, V. (2018). Can very small font size enhance memory? *Memory & Cognition, 46*, 979-993.

Hargis, M. B., & Castel, A. D. (2018). Younger and older adults' associative memory for medication interactions of varying severity. *Memory, 26*, 1151-1158.

Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis*. Academic Press.

Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 39-61). Oxford University Press.

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.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1133-1145.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationship between monitoring and control in metacognition: Lessons for the cause- and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135,* 36-69.

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*, 787-794.

Kuhlmann, B. G., Frank, D. J., & Danner, D. (2020). Louder = larger = clearer: Examining the consistency of metamemory illusions. *Zeitschrift für Psychologie, 228*, 296-300.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.

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*, 1626-1636.

Madan, C. R. (2017). Motivated cognition: Effects of reward, emotion, and other motivational factors across a variety of cognitive domains. *Collabra: Psychology, 3*, 24.

Masson, M. E. J., & 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.

Mueller, M. L., & Dunlosky, J. (2016). How beliefs can impact judgments of learning: analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258.

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical items on people's metamemory judgments: What is the contribution of processing fluency? *Quarterly Journal of Experimental Psychology, 69*, 781-799.

Murphy, D. H., & Castel, A. D. (2020). Responsible remembering: How metacognition impacts adaptive selective memory. *Zeitschrift für Psychologie, 228*, 301-303.

Murphy, D. H., & Castel, A. D. (2021a). Responsible remembering and forgetting as contributors to memory for important information. *Memory & Cognition, 49*, 895-911.

Murphy, D. H., & Castel, A. D. (2021b). Metamemory that matters: Judgments of importance can engage responsible remembering. *Memory, 29*, 271-283.

Murphy, D. H., Agadzhanyan, K., Whatley, M. C., & Castel, A. D. (2021). Metacognition and fluid intelligence in value-directed remembering. *Metacognition and Learning*.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102-116.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26*, 125-141.

Nguyen, L. T., Marini, F., Zacharczuk, L., Llano, D. A., & Mudar, R. A. (2019). Theta and alpha band oscillations during value-directed strategic processing. *Behavioural Brain Research, 367*, 210-214.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12*, 237-241.

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*, 366-388.

Reber, R., & Greifeneder, R. (2017). Perceptual processing fluency in education: How metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist, 52*, 84-103.

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

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.

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

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1236-1242.

Soderstrom, N. C., & Rhodes, M. G. (2014). Metacognitive illusions can be reduced by monitoring recollection during study. *Journal of Cognitive Psychology, 26*, 118-126.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204-221.

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review, 18*, 973-978.

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,* 1024-1037.

Tiede, H. L., & Leboe, J. P. (2009). Metamemory judgments and the benefits of repeated study: Improving recall predictions through

the activation of appropriate knowledge. *Journal of Experimental Psychology: Learning Memory & Cognition, 35*, 822-828.

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, *73*, 629-642.

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, *46*, 507-519.

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S. T., & Shanks, D. R. (2021). How to assess the contributions of perceptual processing fluency and beliefs to the formation of judgments of learning: methods and pitfalls. *Metacognition and Learning*.

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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.