BRIEF REPORT

Judgments of learning reactively facilitate visual memory by enhancing learning engagement

Aike Shi[1] · Chenyuqi Xu[1] · Wenbo Zhao[2] · David R. Shanks[3] · Xiao Hu[4,5] · Liang Luo[1,2] · Chunliang Yang[1,5]

Abstract

Recent studies have found that making judgments of learning (JOLs) for verbal materials changes memory itself, a form of *reactivity* effect on memory. The current study explores the reactivity effect on visual (image) memory and tests the potential role of enhanced learning engagement in this effect. Experiment 1 employed object image pairs as stimuli and observed a positive reactivity effect on memory for visual details. Experiment 2 conceptually replicated this positive reactivity effect using pairs of scene images. Experiment 3 introduced mind wandering (MW) probes to measure participants' attentional state (learning engagement) and observed that making JOLs significantly reduced MW. More importantly, reduced MW mediated the reactivity effect. Lastly, Experiment 4 found that a manipulation that heightened learning motivation decreased the reactivity effect. Overall, the current study provides the first demonstration of the reactivity effect on visual memory, as well as support for the enhanced learning engagement explanation. Practical implications are discussed.

Over the past half century, numerous studies have employed judgments of learning (JOLs; metacognitive estimates about the likelihood of remembering studied materials on a future test) as a measurement tool to assess people's metacognitive awareness about their memory (for reviews, see Rhodes, 2016; Yang et al., 2021). Additionally, it has been well-documented that learners' regulation of their study activities (e.g., decisions about when, what, and how to study) is intimately related to their JOLs (Finn, 2008; Yang et al., 2017). Even though JOLs have long been used as a measure of metacognition, an emerging body of research has established that the act of making a JOL can reactively change the very entity being judged (i.e., memory itself), a phenomenon termed the memory *reactivity* effect (Double et al., 2018; Li et al., 2021; Mitchum et al., 2016; Witherby & Tauber, 2017; Zhao, Li, Shanks, Li, et al., 2022a; Zhao, Li, Shanks, Zhao, et al., 2022b). Below we briefly summarize previous findings about this phenomenon and then introduce the aims and rationale of the current study.

Spellman and Bjork (1992) were the first to speculate that the overt requirement of making JOLs might induce inferential processes which are likely to influence learning itself (Koriat, 1997) and reactively impact memory retention. This hypothesis has subsequently been verified by many studies showing that soliciting JOLs can indeed change memory itself (e.g., Janes et al., 2018; Soderstrom et al., 2015; Zhao, Li, Shanks, Zhao, et al., 2022b). For instance, Soderstrom et al. (2015) asked two (JOL vs. no-JOL) groups of participants to study a list of related word pairs (e.g., *doctor–nurse*). Participants in the JOL group made a JOL when studying each pair, while those in the no-JOL group did not. In a subsequent cued recall test, test performance was better in the JOL than in the no-JOL

Aike Shi and Chenyuqi Xu contributed equally to the present study

✉ Chunliang Yang
chunliang.yang@bnu.edu.cn

[1] Institute of Developmental Psychology, Faculty of Psychology, Beijing Normal University, 19 Xinjiekouwai Street, Haidian District, Beijing 100875, China

[2] Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, China

[3] Division of Psychology and Language Sciences, University College London, London, UK

[4] Faculty of Psychology, Beijing Normal University, Beijing, China

[5] Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Beijing Normal University, Beijing, China

Springer

group, demonstrating a positive reactivity effect (for related findings, see Li et al., 2021; Witherby & Tauber, 2017; Zhao, Li, Shanks, Zhao, et al., 2022b).

Noteworthy is that other studies found that the reactivity effect can be negative under certain conditions (Mitchum et al., 2016). For instance, Mitchum et al. (2016) observed that making JOLs significantly reduced recall of unrelated word pairs, when related and unrelated pairs were studied in a mixed list. Some studies moreover have reported null effects. For example, Ariel et al. (2021) found that making JOLs fails to alter memory for text passages. These findings suggest that the reactivity effect may be moderated by material type. Indeed, a recent meta-analysis found a positive reactivity effect on memory for related word pairs and word lists, while there was minimal reactive influence on memory for unrelated word pairs (Double et al., 2018).

The above discussion leads to an important question: Does the reactivity effect generalize to other types of materials, if it is moderated by material type? Even though studies such as those described above have investigated the reactivity effect on memory for varying types of verbal materials (e.g., word lists, related and unrelated word pairs, text passages), it is an open question whether the effect generalizes to memory for nonverbal materials (e.g., complex visual information). The importance of visual information in daily life and educational settings motivates the current study to examine the reactivity effect on visual memory.

Unlike our limited ability to memorize verbal materials, our capacity for storing visual information is immense (Brady et al., 2008). Investigating the reactivity effect on visual memory provides an opportunity to determine whether the effect is domain specific (i.e., limited to memory for verbal materials) or domain general (i.e., generalizable to memory for other types of information).

Another reason for investigating the reactivity effect on visual memory is that many previous studies elicited item-by-item JOLs to measure people's ability to monitor their visual memory (e.g., Besken, 2016; Undorf et al., 2017). But if making JOLs reactively changes visual memory, and perhaps does so to an extent that is not exactly identical across items, then these JOLs would be inaccurate and contaminated measures of metacognitive ability. Hence, it is important to determine whether making JOLs reactively changes visual memory, and the documented findings might provide guidance for future research design and data interpretation (see General Discussion for details).

Besides exploring the reactivity effect on visual memory, the current study also aims to test the *enhanced learning engagement* (ELE) theory of positive reactivity, which was recently proposed by Zhao, Li, Shanks, Zhao, et al. (2022b) and has thus far only received a small amount of empirical scrutiny. The ELE theory hypothesizes that positive reactivity results from enhanced learning engagement (e.g., study time, attention, effort) induced by the requirement of making JOLs. Specifically, people's attention typically wanes across a prolonged learning episode, resulting in poor learning engagement and more mind wandering (MW), which are harmful for learning and memory (Seli et al., 2018). Making item-by-item JOLs requires participants to focus their mind on the learning task. That is, they have to closely encode and analyze the study items in order to make a reasonable JOL for each of them. Enhanced learning engagement in turn leads to a positive reactivity effect.

Tauber and Witherby (2019) proposed a similar explanation to account for their age difference findings: Making JOLs only reactively enhances cued recall of related word pairs for young but not for older adults. They proposed that the age difference of reactivity might result from the fact that older adults are generally more motivated and their minds typically wander less frequently. Hence, making JOLs is less beneficial for older adults' memory. Although Tauber and Witherby (2019) proposed this explanation, they did not test it. Tauber and Witherby (2019) also proposed several other explanations, which can also readily account for their age difference findings.

Besides Zhao, Li, Shanks, Zhao, et al. (2022b) and Tauber and Witherby (2019), some other researchers have also claimed that making JOLs can enhance learning engagement. For instance, several previous studies asked participants to make JOLs to sustain their attention across a learning task, even though JOLs themselves were not relevant to the primary research questions (e.g., Carpenter & Schacter, 2018). To our knowledge, it has never been directly tested whether making JOLs does indeed maintain learning engagement. Furthermore, it remains unknown whether the ELE theory is a valid explanation of the reactivity effect because this theory has yet to be subjected to empirical tests. Therefore, the second goal of the current study is to test whether enhanced learning engagement is responsible for positive reactivity.

In summary, the current study addresses two important questions regarding reactivity: (1) Whether the reactivity effect generalizes to visual memory, and (2) whether the ELE theory is a valid explanation of the positive reactivity effect. The first question was explored in Experiments 1 and 2, in which participants were instructed to either remember object (Experiment 1) or scene (Experiment 2) images, with half the images studied with concurrent JOLs and the other half without. To foreshadow, both experiments observed strong evidence of a positive reactivity effect on visual memory. The second question was explored in Experiments 3 and 4. Specifically, Experiment 3 employed MW probes to measure participants' learning engagement, and Experiment 4 directly manipulated participants' learning motivation.

Fig. 1 **A** Examples of object image pairs used in Experiment 1. **B** Examples of scene image pairs used in Experiments 2–4

Experiment 1

Experiment 1 was conducted to explore whether making concurrent JOLs reactively changes visual memory.

Method

Participants

A pilot study (with 10 participants) detected a medium-sized (Cohen's $d = 0.53$) reactivity effect on visual memory. A power analysis, conducted via G*Power (Faul et al., 2007), indicated that 30 participants were required to observe a significant (two-tailed, $\alpha = .05$) reactivity effect at 0.80 power. Accordingly, 30 participants ($M_{age} = 20.01$, $SD = 1.84$; 26 female) were recruited from Beijing Normal University (BNU). All of them reported normal or corrected-to-normal vision, did not suffer from memory-related diseases, provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

All experiments reported in the current article were approved by the Ethics Committee of BNU Faculty of Psychology.

Materials

Two hundred object image pairs were selected from the database developed by Brady et al. (2008). Another 10 image pairs were employed for practice. All images were resized to 256 pixels × 256 pixels. As shown in Fig. 1A, the two images for each pair depicted the same object, and only differed minimally in visual details.

To avoid any item-selection effects, for each participant, the program randomly selected one image from each pair to be presented during the study phase, and these images also served as old images presented in the forced-choice recognition test, with their paired counterparts serving as new items. For the 200 to-be-studied images, the program randomly divided them into four blocks, with 50 images in each block. Then the program randomly assigned two blocks to the JOL condition and the other two to the no-JOL condition. The presentation sequence of images in each block and the block sequence were randomized for each participant. All stimuli were presented via MATLAB 2020b Psychtoolbox (Kleiner et al., 2007).

Note that, even though Experiment 1 used images depicting concrete objects, the forced-choice recognition test especially assessed memory for visual details of the studied
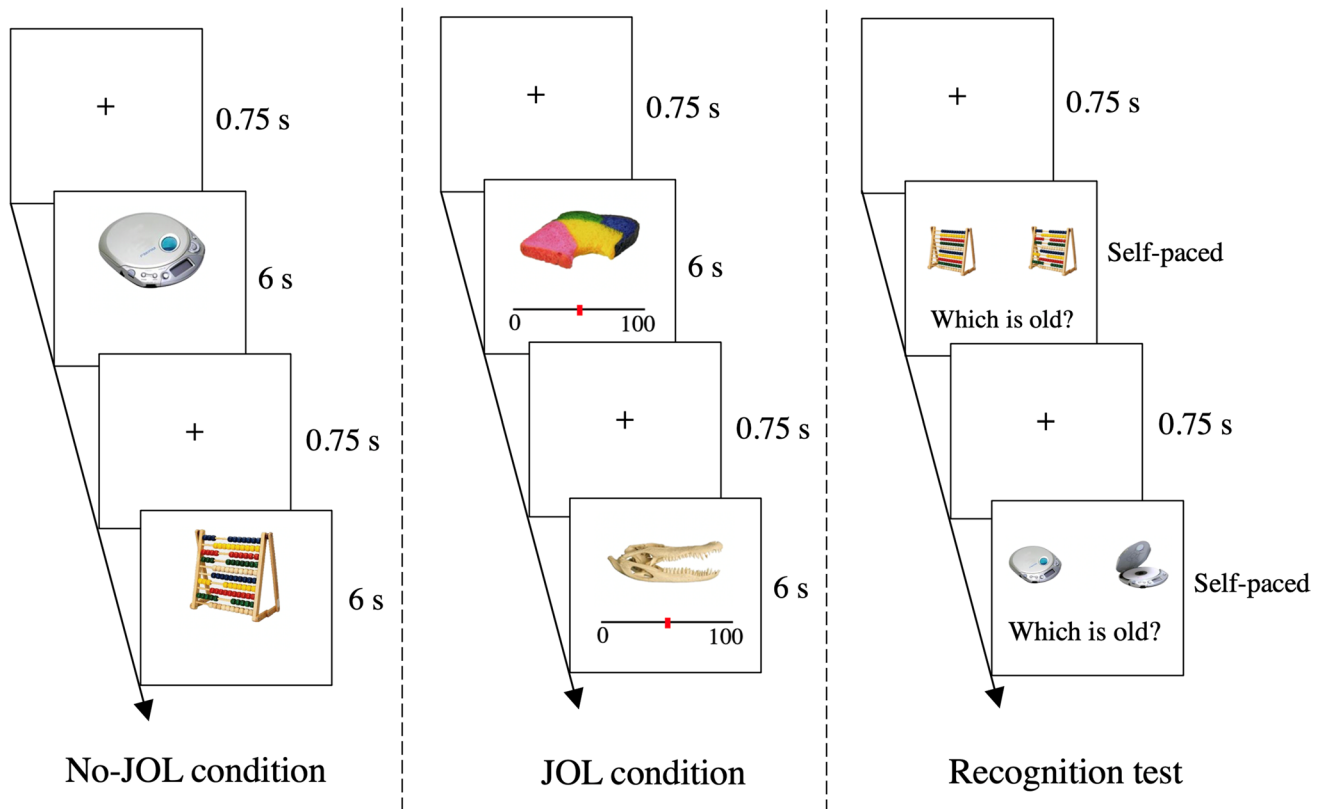
Fig. 2 The left and middle panels depict the task procedure during the study phase in the no-JOL and JOL conditions, respectively. The right panel depicts the task procedure in the forced-choice recognition test

images. That is, in the recognition test, participants had to identify the studied image from a pair of highly similar ones.

### Experimental design and procedure

The experiment involved a within-subjects (study method: JOL vs. no-JOL) design. Participants were told that they would study four blocks of images, with each block consisting of 50 images. For two randomly chosen blocks, they would be asked to make predictions about the likelihood of remembering each image on a later test, while they would not need to make such predictions in the other two blocks. Participants were explicitly instructed to try to memorize all images equally well irrespective of whether they needed to make memory predictions or not, because all images would be eventually tested.

The experimental procedure was adopted from Zhao, Li, Shanks, Zhao, et al. (2022b). Before the formal experiment, participants completed a practice task to familiarize them with the procedure, in which they studied and were tested on the 10 practice image pairs. Then, the main experiment began, the task procedure of which is depicted in Fig. 2. Participants studied four (two JOL and two no-JOL) blocks of images, with 50 images in each block. Before presenting each block, the computer informed participants whether they needed to make memory predictions in the subsequent block.

In a no-JOL block, 50 images were presented one by one in a random order. Before presenting each image, a cross sign appeared at the center of the screen for 0.75 s to mark the interstimulus interval, after which an image appeared for 6 s. Then, the next trial started. This cycle repeated until the end of the block, with a new image studied in each cycle.

The procedure in the JOL blocks was similar to that in the no-JOL block, but with one difference. When an image appeared on the screen, a scale slider, ranging from 0 (*sure I will not remember it*) to 100 (*sure I will remember it*), was simultaneously presented below it (see Fig. 2). Participants were asked to drag and click the slider to make a JOL during the 6-s time window. If they failed to make a JOL, a message box appeared to remind them to carefully make predictions during the required time window for subsequent images. If they successfully made a JOL, the image remained on screen for the remainder of the trial to ensure that the total exposure duration for JOL and no-JOL images was equal.

After studying all four blocks, participants engaged in a distractor task for 5 min, in which they solved as many arithmetic problems (e.g., *52 + 27 = ___*) as they could. Then all participants completed a forced-choice recognition test on all four blocks of images (see Fig. 2). In the recognition test, the 200 old–new image pairs were presented one by one in a random order, with a cross sign presented for 0.75 s between
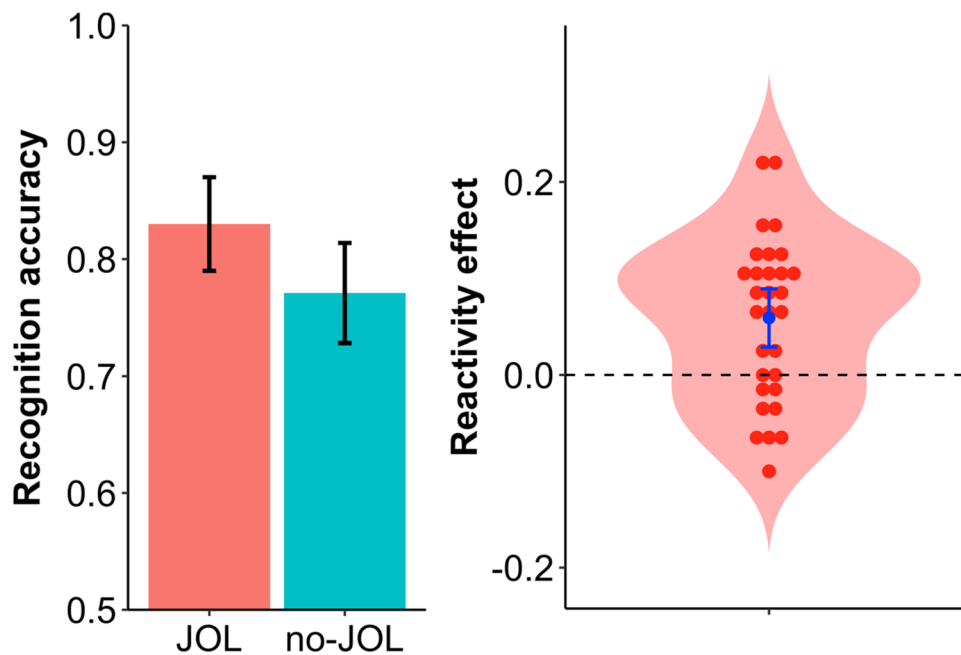
Fig. 3 Recognition accuracy as a function of study method in Experiment 1. In the violin plot (right), each red dot represents a participant's reactivity effect score (i.e., the difference in recognition accuracy between JOL and no-JOL images), with the blue dot representing the group average. Error bars represent 95% CI. (Color figure online)

each two pairs. For each pair, the studied version was randomly presented on the left or right side of the screen. Participants were instructed to decide which image was old (i.e., studied). When a recognition choice was made, the next test trial started automatically. There was no time pressure and no feedback in the forced-choice recognition test.

Results and discussion

Below, we focus on recognition performance (i.e., the reactivity effect). Results regarding item-by-item JOLs are reported in the Supplementary Information (SI). Those results show that participants were underconfident in their judgments, but that JOLs were nonetheless reliably correlated with recognition accuracy (correct choices at test were associated with higher study JOLs than incorrect choices).

As shown in Fig. 3, JOL images ($M = .83$, $SD = .11$) were recognized more accurately than no-JOL ones ($M = .77$, $SD = .12$), difference = .059, 95% CI [.028, .090], $t(29) = 3.85$, $p < .001$, $d = 0.70$, $BF_{10} = 51.42$.[1] As illustrated in the violin plot, a majority (66.7%; 20 out 30) of participants demonstrated positive reactivity, a minority (26.7%) showed negative reactivity, and accuracy in the remaining

two participants (6.7%) was tied (i.e., no reactivity). The proportion showing positive reactivity was substantially larger than the proportion showing negative reactivity, $\chi^2(1) = 8.10$, $p = .004$, and also substantially larger than the proportion showing no reactivity, $\chi^2(1) = 20.74$, $p < .001$.

Overall, these results straightforwardly demonstrate a positive reactivity effect on memory for image details and suggest that the memory reactivity effect is a domain-general phenomenon that is not limited to memory for verbal materials.
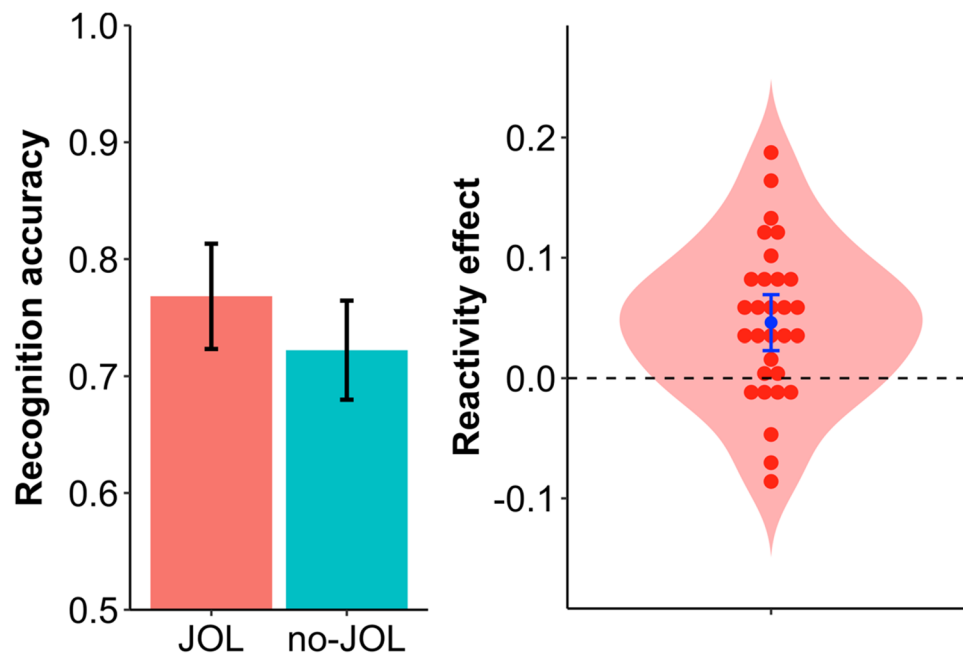
Experiment 2

A possible limitation of Experiment 1 was that the object images were namable, despite the fact that the forced-choice recognition test especially assessed memory for visual details. Experiment 2 was conducted to conceptually replicate the main findings of Experiment 1 by using complex scene images as study stimuli, for which the difference between the two images in each pair (e.g., a change of perspective) had very low nameability (see Fig. 1B).

Method

Participants

A pilot study (with 10 participants) found a medium-sized ($d = 0.57$) reactivity effect on memory for scene images. A power

---

[1] In each of Experiments 1–4, a mixed analysis of variance (ANOVA) was conducted, with block order as a between-subjects variable and study method as a within-subjects variable. The four experiments consistently showed a main effect of study method ($ps \leq .002$), but no main effect of block order ($ps \geq .309$) nor interaction between block order and study method ($ps \geq .136$).

Fig. 4 Recognition accuracy as a function of study method in Experiment 2. In the violin plot (right), each red dot represents a participant's reactivity effect score (i.e., the difference in recognition accuracy between JOL and no-JOL images), with the blue dot representing the group average. Error bars represent 95% CI. (Color figure online)

analysis indicated that 27 participants were needed to detect a significant reactivity effect at 0.80 power. In total, 30 participants ($M_{age} = 21.37$, $SD = 2.21$; 22 female) were recruited from BNU. All of them reported normal or corrected-to-normal vision, did not suffer from memory-related diseases, provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials

Experiment 2 used scene images as stimuli, selected from the database compiled by Konkle et al. (2010). This database consists of a large number of scene images from diverse categories (e.g., airports, amusement parks, shopping malls). From this database, 800 images were selected from 16 categories, with 50 images from each category. All images were resized to 256 pixels × 256 pixels.

Two hundred and fifty-six images, 16 from each of the 16 categories, were randomly selected for each participant for presentation during the study phase. In addition, each of these 256 to-be-studied images was randomly paired with a new image from the same category to form 256 old–new image pairs (see Fig. 1B), which were presented in the forced-choice recognition test. Each studied image was paired with a new image from the same category to increase the difficulty of the forced-choice recognition test.

For each participant, the computer randomly divided the 256 to-be-studied images into four blocks, with each block containing 64 images from four categories. The four blocks were then randomly allocated to the JOL and no-JOL conditions, with two blocks in each condition. The presentation sequence of images in each block and the block sequence were randomly determined for each participant. Another 10 pairs of images were used for practice.

Experimental design and procedure

Apart from the changes noted above, the experimental design and procedure were identical to those in Experiment 1.

Results and discussion

Results of item-by-item JOLs are available in the SI. As before, JOLs were reliably correlated with recognition accuracy.

As shown in Fig. 4, JOL images ($M = .77$, $SD = .13$) were recognized more accurately than no-JOL ones ($M = .72$, $SD = .12$), difference = .046, 95% CI [.022, .071], $t(29) = 3.86$, $p < .001$, $d = 0.71$, $BF_{10} = 53.44$. The proportion (73.3%; 22 out 30) of participants showing positive reactivity was substantially larger than the proportion (23.3%) showing negative reactivity, $\chi^2(1) = 13.08$, $p < .001$, and also substantially larger than the proportion (3.3%) showing no reactivity, $\chi^2(1) = 28.20$, $p < .001$.

These results conceptually replicate the main findings of Experiment 1 by showing a positive reactivity effect on memory for scene images.

Experiment 3

Experiment 3 was designed to test the ELE theory. To achieve this aim, we included MW probes to measure participants' learning engagement in the JOL and no-JOL conditions. According to the ELE theory, which proposes that the requirement to make item-by-item JOLs forces participants to focus more attentively on the learning task, we expected to observe lower MW scores in the JOL than in the no-JOL condition.

Method

Participants

A pilot study (with 10 participants) found a medium-sized (d = 0.53) effect of making JOLs on MW. A power analysis indicated that 30 participants were required to detect a significant effect of making JOLs on MW at 0.80 power. Accordingly, 30 participants (M_{age} = 20.95, SD = 2.16; 28 female) were recruited from BNU. All of them reported normal or corrected-to-normal vision, did not suffer from memory-related diseases, provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design and procedure

The materials, experimental design and procedure were identical to those in Experiment 2, with one exception. Following Peterson and Wissman (2020), Experiment 3 adopted the probe-detection technique to measure participants' learning engagement (indexed by MW scores).

In total, eight MW probes, two in each block, were presented during the study phase. For each participant, the computer randomly presented a MW probe after the presentation of a given image. There were two constraints on the placement of the MW probes. First, two probes were presented in the final three-quarters of each block (in other words, no probes were presented during the first 16 images). Secondly, the two probes in each block were temporally separated by at least five images.

The trial was suspended while the probe was on the screen. The wording of the probe question was as follows: "To what extent were you concentrating on the learning task when you saw this probe? 1 = *I was fully concentrating on the task;* 7 = *I was fully mind-wandering.*" Participants pressed a corresponding number key on the keyboard to respond to each probe. There was no time pressure for responding to the MW probes. MW scores were averaged across the four probes in each of the JOL and no-JOL conditions, with lower MW scores representing greater levels of learning engagement.

Results and discussion

Results relating to the accuracy of item-by-item JOLs are reported in the SI. Once again, JOLs were reliably correlated with recognition accuracy.

As shown in Fig. 5A, JOL images (M = .80, SD = .10) were recognized more accurately than no-JOL ones (M = .74, SD = .12), difference = .056, 95% CI [.029, .082], $t(29)$ = 4.32, p < .001, d = 0.79, BF_{10} = 163.18, replicating the findings from Experiment 2. The proportion (80.0%; 24 out 30) of participants showing positive reactivity was substantially larger than the proportion (20.0%) showing negative reactivity, $\chi^2(1)$ = 19.27, p < .001.

As shown in Fig. 5B, MW scores were lower in the JOL (M = 2.34, SD = 1.07) than in the no-JOL (M = 2.73, SD = 1.31) condition, difference = −0.383, 95% CI [−0.663, −0.104], $t(29)$ = -2.81, p = .009, d = −0.51, BF_{10} = 5.00. The proportion (63.3%; 19 out 30) of participants showing lower levels of MW in the JOL than in the no-JOL condition was substantially larger than the proportion (20.0%) showing the converse pattern, $\chi^2(1)$ = 9.87, p = .002, and also substantially larger than the proportion (16.7%) showing equal levels of MW between the two conditions, $\chi^2(1)$ = 11.74, p < .001.

Furthermore, as shown in Fig. 5C, the difference in MW scores between the JOL and no-JOL conditions was strongly related to the magnitude of the reactivity effect (represented as the difference in recognition performance between the JOL and no-JOL conditions), r = −.58, p < .001, BF_{10} = 43.66, indicating that the more effectively making JOLs reduced MW, the larger the reactivity effect was.

To further explore the relationships among making JOLs (*vs.* not making JOLs), MW, and recognition accuracy, a within-subjects mediation analysis was conducted via the SPSS *MEMORE* package (Montoya & Hayes, 2017). As shown in Fig. 6, the indirect effect of making JOLs on visual memory by reducing MW was significant, $a*b$ = 0.020, 95% CI [0.002, 0.047], suggesting that the reactivity effect on visual memory was at least partially mediated by enhanced learning engagement. The direct effect of making JOLs on visual memory was also significant, c' = 0.036, 95% CI [0.010, 0.061], suggesting that the reactivity effect persisted when the effect of making JOLs on MW was controlled.

Overall, Experiment 3 replicated the positive reactivity effect on memory for scene images. More importantly, it shows that making JOLs reduces MW, and that this reactivity effect is partially mediated by enhanced learning engagement.
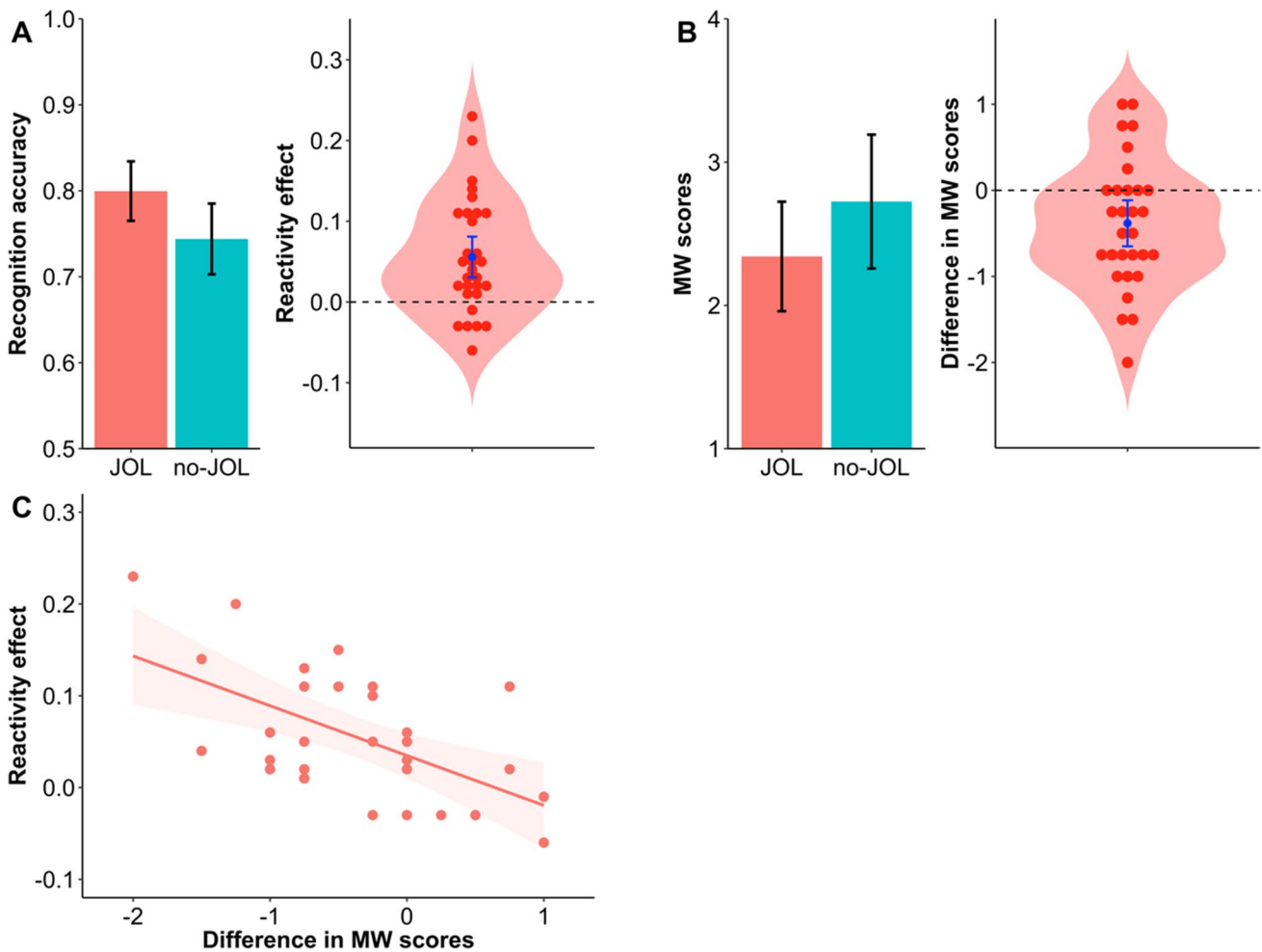
Fig. 5 **A** Recognition accuracy as a function of study method in Experiment 3. In the violin plot (right), each red dot represents a participant's reactivity effect score (i.e., the difference in recognition accuracy between JOL and no-JOL images), with the blue dot representing the group average. **B** Mind wandering (MW) scores as a function of study method. In the violin plot (right), each red dot represents the difference in MW scores between the JOL and no-JOL conditions for a participant, with the blue dot representing the group average. **C** Scatter plot depicting the relationship between the difference in MW scores and the difference in recognition accuracy (i.e., reactivity effect). Each dot shows the data from one participant. Error bars represent 95% CI. (Color figure online)

Experiment 4

A potential limitation of Experiment 3 is that the MW probes might not measure participants' engagement in an unbiased manner. Indeed, these probes might induce a second form of reactivity, themselves changing participants' task performance (Seli et al., 2013; Weinstein et al., 2018; Wiemers & Redick, 2019). Experiment 4 was conducted to further explore the ELE theory in a different way, by directly manipulating participants' learning motivation.
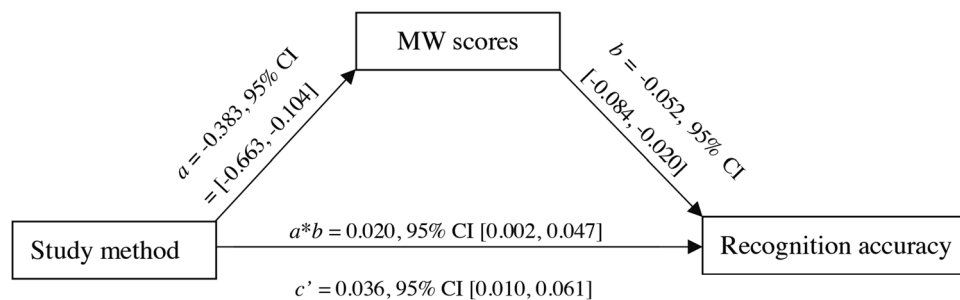
Fig. 6 Mediation results in Experiment 3

Numerous studies have established that learning engagement is positively related to motivation (Guthrie & Cox, 2001), and manipulations that boost motivation can reduce task-based MW (Seli et al., 2019) and improve study effort (Kang & Pashler, 2014). Based on the ELE theory, it is reasonable to expect that enhancing learning motivation would reduce the positive reactivity effect because if participants are highly motivated to perform the learning task well, there would be little room left for JOL-elicitation to further boost learning engagement.

Method

Participants

A pilot study (with 10 participants in each of the motivation and control groups) found that the effect size for the interaction between group (motivation vs. control) and study method (JOL vs. no-JOL) was $\eta_p^2 = .088$. A power analysis indicated that 42 participants in each group were required to detect a significant interaction at 0.80 power. Accordingly, 84 participants ($M_{age} = 21.89$, $SD = 1.77$; 79 female) were recruited from BNU and randomly allocated to the two groups, with 42 in each group. All of them reported normal or corrected-to-normal vision, did not suffer from memory-related diseases, provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design, and procedure

The materials were identical to those in Experiments 2 and 3. The experiment involved a 2 (group: motivation vs. control) $\times$ 2 (study method: JOL vs. no-JOL) mixed design, with group as a between-subjects factor and study method as a within-subjects factor.

Before the learning task, participants in both the motivation and control groups received the same instructions as in Experiment 2. In addition, the motivation group received motivation manipulation instructions adapted from Seli et al. (2019), which were not shown to the control group. The motivation manipulation instructions were as follows:

> As you know, the whole task will take about 1.5 hours to complete. If your memory performance on the final test is lower than the average level observed in our previous study, you will have to spend another 1.5 hours to retake the whole task. This cycle will repeat until your test performance goes above the average level.

After receiving the instructions, participants in both groups completed a practice task, and then started the formal experiment. The procedure of the formal experiment (including study, distractor, and test) was identical to that in Experiment 2.

After completing the recognition test, participants in both groups were instructed to honestly report their motivation to perform well the learning task in order to check whether the motivation manipulation was successful. They were explicitly informed that their motivation reports would not affect them in any way. Participants in the motivation group were further informed that they did not need to retake the task regardless of whether their test performance was above the average level. Their motivation levels were reported on a scale ranging from 1 (*not motivated at all*) to 9 (*very motivated*).

Finally, participants in both groups were informed about the numbers of JOL and no-JOL images they had correctly recognized in the memory test. Participants showing positive reactivity were asked to explain why making memory predictions enhanced their memory, while those showing negative reactivity explained why making memory predictions impaired their memory. Those showing no reactivity explained why making memory predictions had no impact on their memory.

Results and discussion

Results of item-by-item JOLs are reported in the SI. Again, JOLs were reliably related to recognition accuracy. Participants' explanations about the reactivity effect were collected for exploratory analyses, and the detailed results are reported in the SI.

We first conducted a Bayesian t test to check whether the motivation manipulation was successful. The answer was affirmative: The reported motivation scores were significantly higher in the motivation ($M = 7.17$, $SD = 1.10$) than in the control group ($M = 6.29$, $SD = 1.15$), difference $= 0.881$, 95% CI [0.391, 1.371], $t(82) = 3.58$, $p = .001$, $d = 0.78$, $BF_{10} = 47.37$.

A 2 (group: motivation vs. control) $\times$ 2 (study method: JOL vs. no-JOL) Bayesian mixed analysis of variance (ANOVA) was conducted to explore if boosting motivation reduces the reactivity effect, as predicted by the ELE theory. The Bayesian ANOVA was conducted via JASP (Version 0.16.2), with all parameters set at their default values. As shown in Fig. 7, there was a main effect of group, $F(1, 82) = 27.50$, $p < .001$, $\eta_p^2 = .25$, $BF_{incl} = 1.36e+4$, with superior recognition accuracy in the motivation ($M = .87$, $SD = .08$) than in the control group ($M = .77$, $SD = .11$), reflecting that enhancing motivation boosts learning outcomes. There was also a main effect of study method, $F(1, 82) = 34.00$, $p < .001$, $\eta_p^2 = .29$, $BF_{incl} = 8.90e+4$, with JOL images ($M = .84$, $SD = .10$) recognized more accurately than no-JOL ones ($M = .80$, $SD = .13$),
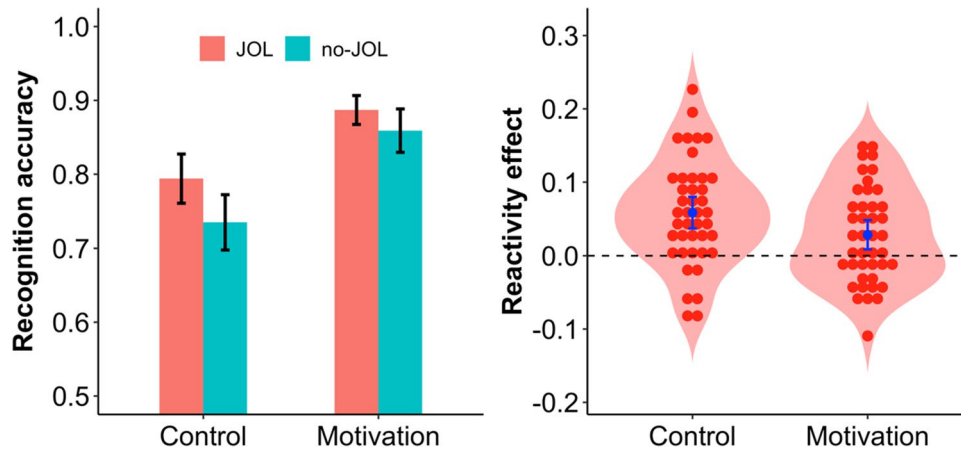
Fig. 7 Recognition accuracy as a function of study method and group in Experiment 4. In the violin plot (right panel), each red dot represents a participant's reactivity effect score (i.e., the difference in rec- ognition accuracy between JOL and no-JOL images), with blue dots representing group averages. Each dot shows the data from one participant. Error bars represent 95% CI. (Color figure online)

reflecting a positive reactivity effect and replicating the results of Experiments 1–3.

Of critical interest, there was a significant interaction between group and study method, $F(1, 82) = 4.04$, $p = .048$, $\eta_p^2 = .05$, $BF_{incl} = 5.09$. This interaction arose from the fact that the positive reactivity effect (calculated as the difference in recognition performance between JOL and no-JOL images) was smaller in the motivation ($M = .028$, $SD = .065$) than in the control group ($M = .058$, $SD = .071$). These results confirm the ELE theory's prediction that a manipulation effective at heightening motivation will reduce the reactivity effect.

In the control group, JOL images ($M = .79$, $SD = .11$) were recognized more accurately than no-JOL ones ($M = .74$, $SD = .12$), difference $= .058$, 95% CI [.036, .082], $t(41) = 5.31$, $p < .001$, $d = 0.82$, $BF_{10} = 4.49e+3$. The proportion (78.6%; 33 out 42) of participants showing positive reactivity was substantially larger than the proportion (14.3%) showing negative reactivity, $\chi^2(1) = 32.36$, $p < .001$, and also substantially larger than the proportion (7.1%) showing no reactivity, $\chi^2(1) = 40.88$, $p < .001$.

In the motivation group, JOL images ($M = .89$, $SD = .07$) were also recognized more accurately than no-JOL ones ($M = .86$, $SD = .10$), difference $= .028$, 95% CI [.008, .049], $t(41) = 2.83$, $p = .007$, $d = 0.44$, $BF_{10} = 5.37$. The proportion (54.8%; 23 out 42) of participants showing positive reactivity was numerically larger than the proportion (38.1%) showing negative reactivity, $\chi^2(1) = 1.72$, $p = .189$, and substantially larger than the proportion (7.1%) showing no reactivity, $\chi^2(1) = 20.11$, $p < .001$. Also confirming the key finding is that the proportion of participants showing positive reactivity was smaller in the motivation (54.8%) than in the control (78.6%) group, $\chi^2(1) = 4.34$, $p = .037$.

Overall, the above results support the ELE theory by showing that enhancing learning motivation reduces the magnitude of the positive reactivity effect and decreases the proportion of individuals whose memory benefits from making JOLs.

General discussion

Previous studies have explored the reactivity effect on memory for verbal materials, such as word lists, related and unrelated word pairs, and text passages (see the Introduction). The current study is the first to investigate whether this effect generalizes to visual memory. Experiments 1–4 consistently found that making concurrent JOLs significantly boosted later recognition accuracy, regardless of whether the study materials were object or scene images. These findings extend the reactivity effect to visual memory and suggest that it is a domain-general phenomenon—although of course extensions to yet other domains, such as auditory memory, will be informative.

Besides extending the reactivity effect to visual memory, the current study also provided the first empirical test of the ELE theory. Specifically, Experiment 3 found that making JOLs significantly enhanced learning engagement (reflected by reduced MW scores), and the level of reduced MW significantly predicted the magnitude of the reactivity effect. Critically, reduced MW partially mediated the reactivity effect. Furthermore, Experiment 4 found that a manipulation which heightened learning motivation reduced the positive reactivity effect and decreased the number of participants showing positive reactivity.

More supporting evidence came from participants' explanations of reactivity, observed in Experiment 4. As shown in the SI, the majority (69.6%) of participants who showed positive reactivity explained that making JOLs facilitated their memory through improving learning engagement. These findings jointly support the ELE theory as a viable

explanatory framework for the reactivity effect (Zhao, Li, Shanks, Zhao, et al., 2022b).

Interestingly, Experiment 3 found that after controlling the effect of making JOLs on MW, the reactivity effect survived. Similarly, Experiment 4 found that enhancing learning motivation reduced but did not eliminate the reactivity effect (that is, the reactivity effect persisted in the motivation group). These findings suggest that the ELE theory does not provide a complete explanation of this effect, and there are other mechanisms through which making JOLs benefits visual memory.

What might such mechanisms be? One possibility is that the requirement to make concurrent JOLs changes the encoding strategies participants employ, as suggested by the strategy-change theory of reactivity (Mitchum et al., 2016). Indeed, as shown in the SI, 41.1% of participants showing positive reactivity explained that making JOLs enhanced their memory because they used better strategies in the JOL condition, such as searching for distinctive features of the images, focusing more on visual details of the images, and self-evaluation. It should be acknowledged that participants' explanations of reactivity were subjective, and more experimental research is required to directly test the role of strategy change in the reactivity effect on visual memory. It should also be noted that the current study only tested the role of enhanced learning engagement in the reactivity effect on visual memory. Future research needs to test the ELE theory's validity in explaining the reactivity effects on memory for other types of materials, such as word lists (Li et al., 2021; Zhao, Li, Shanks, Zhao, et al., 2022b).

Putting the theoretical implications aside, the findings obtained here also bear practical implications for guiding future research design and interpretation. Some previous studies asked participants to make item-by-item JOLs to measure their metamemory accuracy in monitoring visual memory (e.g., Besken, 2016; Undorf et al., 2017). However, Experiments 1–4 consistently showed that making JOLs reactively changed visual memory, highlighting a potential drawback of this procedure: Inferences about the memory-metamemory relationship in a standard no-JOL condition cannot be inferred from the memory-metamemory relationship observed in a JOL condition. Hence, future metamemory research needs to develop more elegant methods to prevent or alleviate this reactivity effect when assessing JOL accuracy. At the very least, researchers should bear the reactivity effect in mind when interpreting their metamemory results.

Some studies have asked participants to make item-by-item JOLs in order to sustain their attention across a learning task (e.g., Carpenter & Schacter, 2018). However, the assumption that making JOLs improves learning engagement has not been verified before. Hence, another contribution of the current study is that it provides direct evidence justifying this assumption. The corresponding practical implication is that making item-by-item JOLs can be applied as a practice to maintain learning engagement across a prolonged learning episode.

In conclusion, making metamemory judgments (JOLs) enhances learning engagement and reactively boosts visual memory. The ELE theory is a viable explanation for the reactivity effect on visual memory.

References

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review, 33*, 693–712.

Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1417–1433.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*, 14325–14329.

Carpenter, A. C., & Schacter, D. L. (2018). False memories, false preferences: Flexible retrieval mechanisms supporting successful inference bias novel decisions. *Journal of Experimental Psychology: General, 147*, 988–1004.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgments of learning. *Memory, 26*, 741–750.

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

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition, 36*, 813–821.

Guthrie, J. T., & Cox, K. E. (2001). Classroom conditions for motivation and engagement in reading. *Educational Psychology Review, 13*, 283–302.

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review, 28*, 2356–2364.

Kang, S. H. K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition, 3*, 183–188.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. *Perception, 36*(16, ECVP Abstract Supplement), 1–6.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*, 1551–1556.

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.

Li, B., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., Yin, Y., Liu, M., Yang, C., & Luo, L. (2021). Soliciting judgments of forgetting reactively enhances memory as well as making judgments of

learning: Empirical and meta-analytic tests. *Memory & Cognition*. Advance online publication. https://doi.org/10.3758/s13421-021-01258-y

MATLAB. (2020). *version 9.9.0 (R2020b)*. The MathWorks Inc.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145*, 200–219.

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*, 6–27.

Peterson, D. J., & Wissman, K. (2020). Using tests to reduce mind-wandering during learning review. *Memory, 28*, 582–587.

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

Seli, P., Carriere, J., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology, 4*, 430.

Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018). Mind-wandering as a natural kind: A family-resemblances view. *Trends in Cognitive Sciences, 22*, 479–490.

Seli, P., Schacter, D. L., Risko, E. F., & Smilek, D. (2019). Increasing participant motivation reduces rates of intentional and unintentional mind wandering. *Psychological Research, 83*, 1057–1069.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 553–558.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3*, 315–317.

Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology & Aging, 34*(6), 836–847.

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language, 92*, 293–304.

Weinstein, Y., De Lima, H. J., & van der Zee, T. (2018). Are you mind-wandering, or is your mind on task? The effect of probe framing on mind-wandering reports. *Psychonomic Bulletin & Review, 25*, 754–760.

Wiemers, E. A., & Redick, T. S. (2019). The influence of thought probes on performance: Does the mind wander more if you ask it? *Psychonomic Bulletin & Review, 26*, 367–373.

Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition, 6*, 496–503.

Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1073–1092.

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: Methods and pitfalls. *Metacognition and Learning, 16*, 319–343.

Zhao, W. B., Li, J., Shanks, D. R., Li, B., Hu, X., Yang, C., & Luo, L. (2022a). Metamemory judgments have dissociable reactivity effects on item and inter-item relational memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Advance online publication. https://doi.org/10.1037/xlm0001160

Zhao, W. L., Li, B., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., Yin, Y., Luo, L., & Yang, C. (2022b). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children's learning. *Child Development, 93*, 405–417.