Metacognition and Learning
https://doi.org/10.1007/s11409-024-09382-1

Soliciting judgments of learning reactively facilitates both recollection- and familiarity-based recognition memory

Jun Zheng[1,2] · Baike Li[3] · Wenbo Zhao[4] · Ningxin Su[2] · Tian Fan[2] · Yue Yin[2] · Yali Hu[2] · Xiao Hu[5] · Chunliang Yang[5,6] · Liang Luo[6,7]

Received: 22 January 2024 / Accepted: 3 March 2024

Abstract
Successful recognition is generally thought to be based on both recollection and familiarity of studied information. Recent studies found that making judgments of learning (JOLs) can reactively facilitate recognition performance, a form of *reactivity effect* on memory. The current study aimed to explore the roles of recollection and familiarity in the reactivity effect on recognition performance. Experiment 1 replicated the positive reactivity effect on recognition performance. Experiment 2 used the sequential remember/know (R/K) procedure, Experiment 3 utilized the simultaneous R/K procedure, and Experiment 4 inserted a long study-test interval (i.e., 24-h) to determine the roles of recollection and familiarity in the reactivity effect. These three experiments converged in demonstrating that making JOLs reactively facilitated recognition performance through enhancing both recollection and familiarity. Furthermore, there was minimal difference between the reactive influences on recollection and familiarity. The documented findings imply that the JOL reactivity effect on recognition is supported by two underlying mechanisms: greater recollection induced by enhanced distinctiveness, and superior familiarity induced by enhanced learning engagement.

Keywords Judgments of learning · Reactivity · Recognition · Recollection · Familiarity

Introduction

Judgments of learning (JOLs) refers to people's predictions about the likelihood of remembering a studied item in a future memory test (Nelson & Dunlosky, 1991). Over the past few decades, numerous studies have employed JOLs to measure people's metacognitive monitoring (e.g., Besken & Mulligan, 2013; Hu et al., 2015; Rhodes & Castel, 2008; Yang et al., 2018), which is essential for learners to effectively regulate their learning activities (Bjork et al., 2013; Thiede et al., 2003). Researchers have long been aware of the potential reactive influence of measuring the ongoing metacognitive monitoring process on the cognitive process itself (Fox et al., 2011; Spellman & Bjork, 1992). However, many studies presumed that JOLs are a neutral measurement tool for assessing people's metacognitive awareness

Extended author information available on the last page of the article

Published online: 27 March 2024

of their memory status (Nelson & Narens, 1990). Recently, a growing body of research has documented that the act of making JOLs can reactively alter the memory processes being monitored. That is, soliciting JOLs produces a *reactivity effect on memory* (Double et al., 2018; Li et al., 2022; Mitchum et al., 2016; Shi et al., 2022; Soderstrom et al., 2015; Zhao et al., 2022).

Most of previous studies used word pairs as learning materials and examined the reactive influence of making JOLs on cued recall performance (e.g., Janes et al., 2018; Li et al., 2023; Mitchum et al., 2016; Rivers et al., 2021; Soderstrom et al., 2015; Witherby & Tauber, 2017). The results consistently found a positive reactivity effect on cued recall of related word pairs. For instance, Soderstrom et al. (2015) instructed two groups of participants to study strongly related and weakly related word pairs for a later memory test. In a JOL group, participants studied a pair for 8 s and were required to make a JOL during the second half of the exposure duration (i.e., the last 4 s). Whereas, in a no-JOL group, participants studied each pair for 8 s in total without making JOLs. Following the study phase and 3-min distraction task, participants were asked to complete a cued recall test. The results showed that the JOL group recalled significantly more strongly related pairs than the no-JOL group. The positive reactivity effect on memory for related word pairs remains to be very strong even after a long (i.e., 2 days) retention interval (Witherby & Tauber, 2017).

By contrast to cued recall test, another important test format, that is, recognition test, has received less attention (Li et al., 2022; Zhao et al., 2022). Unlike cued recall tests, recognition tests require participants to correctly identify previously studied items among a set of studied and new items (Shepard, 1967; Yonelinas, 2002). When learning a list of unrelated words, participants need to focus more on the specific characteristics of the study items, rather than the association between the cue and target words (Hockley & Consoli, 1999). A large number of cognitive, neuropsychological, and neuroimaging studies have demonstrated that recognition involves two distinct memory processes (or components): recollection and familiarity (Curran & Hancock, 2007; Eichenbaum et al., 2007; Skinner & Fernandes, 2007; Yonelinas, 2002). Specifically, recognition memory is generally thought to be based on both recollection of rich contextual details about previous events or on assessment of stimulus familiarity without recollecting such contextual details.

Previous studies have demonstrated that the acting of making JOLs can enhance recognition of word lists (Li et al., 2022; Tekin & Roediger, 2020; Zhao et al., 2022). However, an important question yet to be addressed with the reactivity effect on recognition memory concerns whether and to what extent making JOLs benefits different components (i.e., recollection and/or familiarity) of recognition memory. Put differently, it has never been explored whether making JOLs reactively strengthens recollection-based recognition memory, familiarity-based recognition memory, or both. Thus, the main aim of the current study was to explore this critical question.

Below we briefly summarize previous empirical findings that inform the hypotheses of the current study, then discuss the potential mechanisms underlying the JOL reactivity effect on recognition memory, and finally provide an overview of the current study.

Empirical findings on recollection and familiarity

Numerous studies have established that recollection and familiarity are two independent memory components that support recognition memory together. Behavioral experiments have shown that some manipulations affect only one component but not the other (for reviews, see Eichenbaum et al., 2007; Yonelinas, 2002). For instance, Cohen et al. (2017) utilized the remember/know (R/K) procedure, which is one of the most widely-used approaches to assess recollection and familiarity (Eichenbaum et al., 2007; Tulving, 1985), to measure the proportion of recollection-based recognition and familiarity-based recognition in value-directed memory. In the R/K test, participants were first instructed to judge whether an item was "old" or "new". Then, if an item was identified as "old", they also need to report whether their "old" decision was based on "remember" (indicating recollection) or "know" (indicating familiarity). The results of Cohen et al. (2017) showed that estimated recollection was greater for high- than for low-value items. However, there was no difference in estimated familiarity between the high- and low-value items. These findings suggest that value improves recognition memory mainly through enhancing the recollection component but not affecting the familiarity component. However, it should be highlighted that some other manipulations, such as rote rehearsal (Dobbins et al., 2004), have been found to only affect familiarity, not recollection.

Recollection and familiarity also show dissociation in brain neural activities (for reviews, see Eichenbaum et al., 2007; Rugg & Curran, 2007). ERP research demonstrated that a 300- to 500-ms mid-frontal FN400 is related to familiarity (Addante et al., 2012; Bader & Mecklinger, 2017; Curran & Hancock, 2007), whereas a 500- to 800-ms parietal component, the LPC, is associated with recollection (Curran, 2004; Vilberg et al., 2006; Woodruff et al., 2006). Evidence from neuroimaging and patients with brain injuries again demonstrated the distinct effects of recollection and familiarity on recognition memory (Aggleton et al., 2005; Bastin et al., 2019; Brown & Aggleton, 2001; Staresina et al., 2012).

Returning to JOL reactivity, several studies have investigated the impact of JOLs on recognition memory for word lists (Li et al., 2022; Tekin & Roediger, 2020; Zhao et al., 2022). For instance, Li et al. (2022) instructed participants to study four lists of unrelated words, with two lists studied with item-by-item JOLs and the other two without. The old/new recognition test results showed superior recognition performance for JOL than for no-JOL words, reflecting a positive reactivity effect on recognition memory. Recognition memory is crucial for individuals across various age populations, from young children learning vocabulary to university students memorizing the specialized terminologies, and to older adults memorizing the items in a shopping list. Elucidating the specific impacts of making JOLs on recognition process (i.e., recollection and familiarity) is not only an important step to understand the mechanism underlying JOL reactivity but also holds important practical implications.

Examining the reactive influences of making JOLs on recollection and familiarity could offer insights into previous research findings. For instance, Myers et al. (2020) observed that test format moderated the JOL reactivity effect, with the JOL reactivity effect being significantly larger on recognition tests than on free recall tests. A potential explanation is that soliciting JOLs may enhance both recollection and familiarity of studied items. As previously mentioned, both recollection and familiarity contribute to successful recognition, whereas free recall primarily relies on recollection (Yonelinas, 2002). Hence, the larger reactivity effect on recognition may stem from the additional enhancement effect of JOLs on familiarity (Zhao, Li et al., 2023a; Zhao, Yin et al., 2023b). Although this explanation sounds reasonable, it should be highlighted that this explanation has not been subjected to empirical tests.

Prior literature has shown that several encoding manipulations, which direct increased attention to the characteristics (e.g., meaning, semantic information) of study materials, can simultaneously enhance both recollection and familiarity (Yonelinas, 2002). These encoding manipulations include asking participants to engage in a deep versus a shallow level of processing (e.g., Craik & Lockhart, 1972; Sheridan & Reingold, 2012), requiring participants to generate rather than passively read a study item (e.g., Sheridan & Reingold, 2011; Slamecka & Graf, 1978), or instructing participants to read words aloud instead of silently reading them (e.g., Fawcett & Ozubko, 2016; Ozubko et al., 2012). Additionally, divided attention (e.g., distracting learning by asking participants to press a key each time three odd digits occurred consecutively) diminishes both recollection and familiarity (Curran, 2004; Gruppuso et al., 1997; Mangels et al., 2001; Yonelinas, 2001).

Taking the study conducted by Ozubko et al. (2012) as a concrete example, in which participants were instructed to study a list of 80 words, reading half of them aloud and the other half silently. They employed R/K procedure to assess recollection and familiarity, in which participants were asked to make a "new", "know", or "remember" response for each test word. The results showed that the proportions of words correctly recognized based on recollection and familiarity were significantly higher for aloud than for silent words. In other words, production (i.e., reading words aloud) increased both recollection and familiarity. Similar to the aforementioned encoding operations, the requirement of making item-by-item JOLs forces participants to devote more attention to the mnemonic characteristics of the learning items (Shi et al., 2022; Zhao et al., 2022). Hence, this increased engagement should produce better retrieval of episodic details and greater familiarity of JOL words. Thus, we hypothesize that making JOL can reactively enhance both recollection and familiarity.

Mechanisms underlying JOL reactivity

The *cue-strengthening theory* (Soderstrom et al., 2015) and *changed-goal theory* (Mitchum et al., 2016) are two hypotheses proposed to account for the JOL reactivity effect on cued recall of word pairs. The former theory posits that the act of making JOLs strengthens the semantic association between the cue and target words, thereby enhancing cued recall of related word pairs. This theory accounts for why soliciting JOLs can improve cued recall of related word pairs and also predicts a lack of reactivity effect on cued recall of unrelated word pairs (e.g., Myers et al., 2020; Rivers et al., 2021; Witherby & Tauber, 2017). The latter theory suggests that when learning a mixed list of difficult (e.g., unrelated word pairs) and easy (e.g., related word pairs) items, making JOLs heightens the learners' metacognitive awareness of the differences in learning difficulty, consequently altering the learning goal towards mastering the easy items meanwhile sacrificing the difficult ones. This theory explains why making JOLs amplifies the difference in cued recall performance between related and unrelated word pairs (Janes et al., 2018; Mitchum et al., 2016). Given that cue-strengthening theory (Soderstrom et al., 2015) and changed-goal theory (Mitchum et al., 2016) are mainly proposed to account for the JOL reactivity effect on cued recall of word pairs whereas the current study targets to explore the JOL reactivity effect on recognition of word lists, we do not further discuss these two theories below.

A recently proposed theory, that is, the *enhanced learning engagement (ELE) theory* (Shi et al., 2022; Zhao et al., 2022), hypothesizes that the positive reactivity effect results from enhanced learning engagement (e.g., attention and cognitive effort) induced by the requirement of making JOLs. This theory suggests that during a prolonged learning task, participants' attention to the task gradually wanes, accompanied by an increase of attentional dispersion behaviors such as mind wandering. Making item-by-item JOLs requires participants to carefully analyze each study item, find appropriate cues to guide JOL formation, and then provide a reasonable JOL for each item. Thus, the enhanced learning engagement in turn leads to a positive reactivity effect. As suggested by this theory and corroborated by related empirical evidence previously introduced, the positive JOL reactivity effect on recognition memory may derive from both enhanced recollection and enhanced familiarity.

More specifically, through augmenting learning engagement, making JOLs directs greater focus to the mnemonic characteristics (e.g., meaning) of study items, which in turn increases the likelihood of recalling more contextual details of the JOL words. On the other hand, individuals may render judgments based on the experience of familiarity that stem from the strength of the memory trace (Yonelinas, 2002). Making JOLs has the potential to sustain attention throughout the learning phase, which contributes to the overall memory strength of the JOL words, ultimately enhancing their familiarity-based recognition.

Overview of the current study

To the best of our knowledge, no prior research has explored the roles of recollection and familiarity in the JOL reactivity effect on recognition memory. Therefore, the primary objective of the current study was to fill this gap. Experiment 1, serving as the groundwork for subsequent experiments, was conducted to replicate the positive JOL reactivity effect on recognition memory, in which half of the words were studied with concurrent JOLs and the other half without. After successful replication of this effect, Experiment 2 utilized sequential R/K procedure (where an old/new judgment precedes an R/K judgment) and Experiment 3 employed the simultaneous R/K procedure (where an R/K/New judgment is rendered within a single trial) to quantify the roles of recollection and familiarity in the JOL reactivity effect on recognition memory. To ensure the robustness of our findings and to assess the durability of the JOL reactivity effects on recollection and familiarity, Experiment 4 extended the study-test interval from 5 min to 24 h. Finally, a meta-analysis was performed to integrate results across the four experiments to increase statistical power (1) to further determine the magnitude of the JOL reactivity effects on recollection and familiarity and (2) to determine whether the effect on recollection is stronger or weaker than the effect on familiarity.

Experiment 1

Experiment 1 was conducted to investigate whether making concurrent JOLs can reactively improve recognition memory for word lists, as a replication of previous findings (Li et al., 2022; Tekin & Roediger, 2020; Zhao et al., 2022), and also to lay the groundwork for further research on the specific reactive influences on the internal processes (i.e., recollection and familiarity) of recognition memory.

Method

Participants

Previous research (Li et al., 2022, Experiment 2) detected a large (Cohen's $d=1.23$) reactivity effect of JOLs on recognition memory for word lists. A power analysis, conducted via G*Power (Faul et al., 2007), suggested a minimum sample size of 8 participants to observe a significant (two-tailed $\alpha=0.05$) reactivity effect at 0.80 power. To be more conservative, we decided to increase the sample size to 25. The final sample consisted of 26 participants ($M_{age}=19.96$, $SD=1.64$; 21 female) recruited from Beijing Normal University (BNU) participant pool. All participants provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation for their participation. All experiments reported in the current study were approved by the Ethics Committee of BNU Faculty of Psychology.

Materials

The materials were 330 two-character Chinese words selected from the Chinese word database developed by Cai and Brysbaert (2010). The word frequency ranged from 1.4 to 20.45 per million ($M_{frequency}=9.73$, $SD=5.45$). The number of strokes ranged from 5 to 35 ($M_{stroke}=17.31$, $SD=4.98$). Ten words were used for practice and the remaining 320 words were used for the formal experiment. To avoid any item-selection effects, for each participant, the program randomly selected half of these 320 words to be presented during the study phase, and these words also served as "old" words in the recognition test, with the other 160 words serving as "new" items. The 160 to-be-studied words were randomly divided into four lists, with 40 words in each list. Two lists were randomly assigned to the JOL condition and the other two to the no-JOL condition. The presentation sequence of words in each list and the list sequence were randomized for each participant. All stimuli were presented via the Matlab 2020b *Psychtoolbox-3* package (Kleiner et al., 2007).

Design and procedure

The experiment involved a within-subjects design (study method: JOL *vs.* no-JOL). Participants were informed that they would study four lists of words, with 40 words in each list, in preparation for a later memory test. For two lists, they would be asked to make predictions about the likelihood of remembering each word in the final test (i.e., JOL condition), and they would not need to make such predictions for the other two lists (i.e., no-JOL condition). Importantly, participants were explicitly instructed to try to memorize all words equally well regardless of whether they needed to make memory prediction or not, because all words would be eventually tested.

The experimental procedure was adapted from Li et al. (2022). Before the formal experiment, participants completed a practice task to familiarize themselves with the task requirements. Then the formal experiment commenced. The formal experiment consisted of three phases: the study phase, the distractor phase, and the final test phase. In the study phase, participants studied four lists of words, with 40 words in each list. Before presenting each list, the computer informed participants whether they would need to make memory predictions for the upcoming list of words.

In a no-JOL list, the 40 words were presented one-by-one in a random order with a 6-s presentation time for each word, preceded by a 500-ms fixation cross (“+”). The procedure for the JOL lists was similar to that for the no-JOL lists, but with one difference. Specifically, after a word presented on-screen for 3 s, a scale slider, ranging from 0 (*sure I will not remember it*) to 100 (*sure I will remember it*), appeared below the word. Participants were asked to drag and click the slider to provide a memory prediction during the last 3 s. The initial position of the mouse was set at the middle of the slider (50), and the slider would display the corresponding value when the mouse moved. If participants failed to make a JOL within the 3-s time window, a message box appeared to remind them to carefully make memory predictions during the required time window for the following words. If they successfully made a JOL, the word and slider remained on screen until the end of the trial, ensuring that the total exposure time (i.e., 6 s) of JOL and no-JOL words was equal.

After the study phase, participants solved as many simple mathematical problems as they could for 5 min. Immediately following the distractor task, participants completed an old/new recognition test. The 160 studied and 160 new words were presented one at a time in a random order, with a 500-ms fixation cross (“+”) presented before each word. Participants were instructed to judge whether the on-screen word was “old” (i.e., studied) or “new” (i.e., unstudied). If it was identified as a new word, keycode “F” should be pressed; if it was identified as an old word, keycode “J” should be pressed. The prompt for the keycodes was always presented below the word. The recognition test was self-paced, and no feedback was provided.

Results and discussion

Below, we focus on recognition performance (i.e., hit rates for both conditions).[1] Results regarding item-by-item JOLs are reported in the Supplementary Information (SI). As a summary, those results showed that participants were overall underconfident in their memory performance, but that JOLs were nonetheless reliably correlated with recognition accuracy (i.e., high JOL words were more likely to be correctly recognized than low JOL ones).

A paired *t*-test showed that recognition accuracy for JOL words ($M=0.84$, $SD=0.12$) was significantly greater than that for no-JOL words ($M=0.67$, $SD=0.21$), difference = 0.16, 95% CI = [0.10, 0.23], $t(25)=5.46$, $p<.001$, $d=1.07$, $BF_{10}=1.92e+3$ (see Fig. 1), replicating the positive JOL reactivity effect on recognition performance. As illustrated in the violin plot, a majority (88.5%; 23 out 26) of participants demonstrated a positive reactivity effect, with only a minority (11.5%) showing a negative reactivity effect. The proportion showing positive reactivity was substantially larger than the proportion showing negative reactivity, $\chi^2(1)=15.39, p<.001$. Consistent with previous studies, these results successfully replicated the positive reactivity effect on recognition memory.

[1] The current study employed a within-subjects design of study method (JOL *vs.* no-JOL), which means that false alarm rates were identical between the JOL and no-JOL conditions. Hence, we mainly took hit rates as the key measure of JOL reactivity. We note that the results of signal-detection *d*’ were identical to those of hit rates.

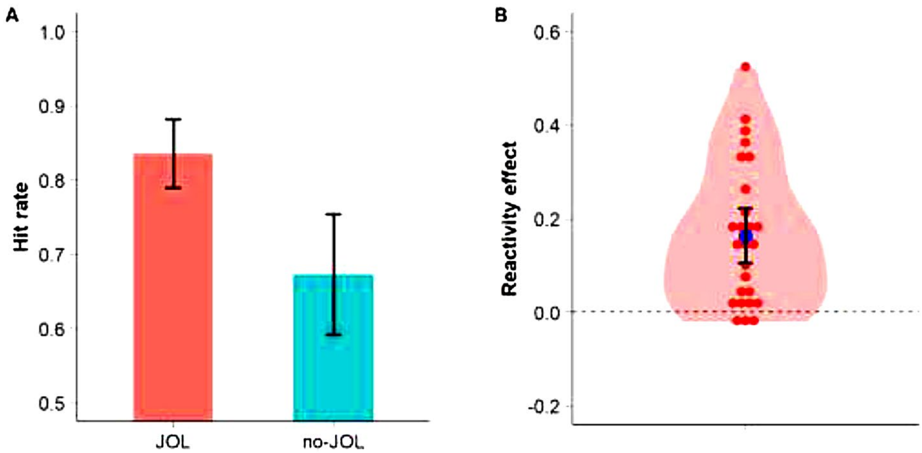

Fig. 1 Results of Experiment 1. Panel A: Hit rates for JOL and no-JOL words in Experiment 1. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in hit rates between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs

Experiment 2

Experiments 2 targeted to explore the reactivity effects on recollection and familiarity by using the classical remember/know (R/K) procedure, one of the most widely-used methods for assessing recollection and familiarity (Eichenbaum et al., 2007; Tulving, 1985). In the R/K procedure, participants were explicitly required to introspect about the basis of their "old" responses and report whether they recognize each "old" item on the basis of remembering (i.e., recollection of episodic information about the study item) or knowing (i.e., being familiar with the study item but without recollecting contextual details) (Yonelinas, 2002).

Before moving forward, a limitation of the R/K procedure should be elaborated. Specifically, "know" responses do not provide an unbiased measure of familiarity because the proportion of "know" responses is affected by the proportion of "remember" responses (Yonelinas, 2002). That is, in the R/K procedure, participants are only allowed to make a "know" response when an item is "familiar without recollection", rather than whenever they feel familiar with the item regardless of whether any recollection occurs. As the proportion of recollection responses increases, the proportion of knowing responses decreases correspondingly (Brown & Bodner, 2011). Thus, the raw proportion of "know" responses tends to underestimate the probability that an item is recognized based on familiarity.

Researchers have developed an independence remember/know (IRK) method to compensate for this underestimation (Cohen et al., 2017; Jacoby et al., 1997; Ozubko et al., 2012; Rosenstreich & Ruderman, 2017; Yonelinas & Jacoby, 1995). In the IRK method, the proportion of "remember" responses (R) is used directly as an estimate of recollection, and familiarity is measured by dividing the proportion of "know" responses (K) by one minus the proportion of "remember" responses (i.e., $F = K/(1 - R)$). Previous evidence has supported the effectiveness of the IRK method (Mangels et al., 2001; Ozubko et al., 2012; Yonelinas, 2002). Hence, the IRK method was used to calculate the proportions of recollection and familiarity in the current study.

Method

Participants

According to the effect size observed in Experiment 1 ($d = 1.07$), a power analysis showed that at least 10 participants were needed to observe a significant (two-tailed $\alpha = 0.05$) reactivity effect at 0.80 power. To be more conservative and also to detect potential reactivity effects on recollection and familiarity, we decided to increase the sample size to 25. In total, 28 participants ($M_{age} = 21.07$, $SD = 2.34$; 23 female) were recruited from BNU participant pool. All participants provided informed consent, were tested individually in a soundproofed cubicle, and received monetary compensation.

Materials, design and procedure

The materials, experimental design and procedure were identical to those in Experiment 1, but with one exception. The familiarity- and recollection-based processes were assessed using the sequential R/K procedure (Eldridge et al., 2002), in which participants were asked first to provide an old/new response followed by an R/K response. Specifically, in the final test, the 160 studied (old) and 160 new words were presented one-by-one in a random order. Participants were instructed to judge whether the on-screen word was "new" (pressing the "F" key) or "old" (pressing the "J" key). If a "new" response was made, the test trial started automatically. If an "old" response was made, participants had to further identify whether the response is based on "familiar" (pressing the "V" key) or "remember" (pressing the "N" key). The term "familiar" was used instead of the standard term "know" in order to avoid confusion due to vague meaning of the word "know" outside of memory laboratories. There was a 500-ms blank interval between "old/new" and "R/K" judgments, and during the "R/K" judgment the words remained on the screen. The prompts for the keycodes were always displayed below each word. No feedback was provided during the final test.

Participants were asked to make "remember" responses when they were able to consciously recall the details or thoughts they had experienced while learning the word. "Familiar" responses were to be made when recognition of the word was accompanied by feelings of familiarity but without evoking any specific conscious recollection. After explaining the instructions, participants were asked if they had any questions about the task requirements. If anything, the experimenter re-explained the difference between "remember" and "familiar" responses until the participants no longer demonstrated confusion. After the test, participants were interviewed and asked to provide examples of each type of responses. The results showed that all participants understood and followed the task requirements.

Results and discussion

Regarding hit rates, JOL words ($M = 0.88$, $SD = 0.10$) were recognized more accurately than no-JOL words ($M = 0.74$, $SD = 0.17$), difference $= 0.14$, 95% CI = [0.08, 0.19], $t(27) = 5.49$, $p < .001$, $d = 1.04$, $BF_{10} = 2.59e+3$ (see Fig. 2), replicating the positive reactivity effect on

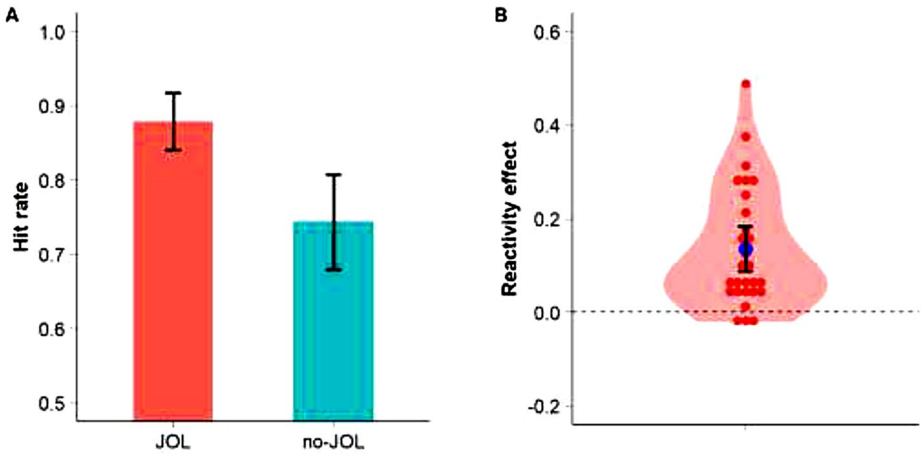

Fig. 2 Hit rate results in Experiment 2. Panel A: Hit rates for JOL and no-JOL words in Experiment 2. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in hit rates between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs

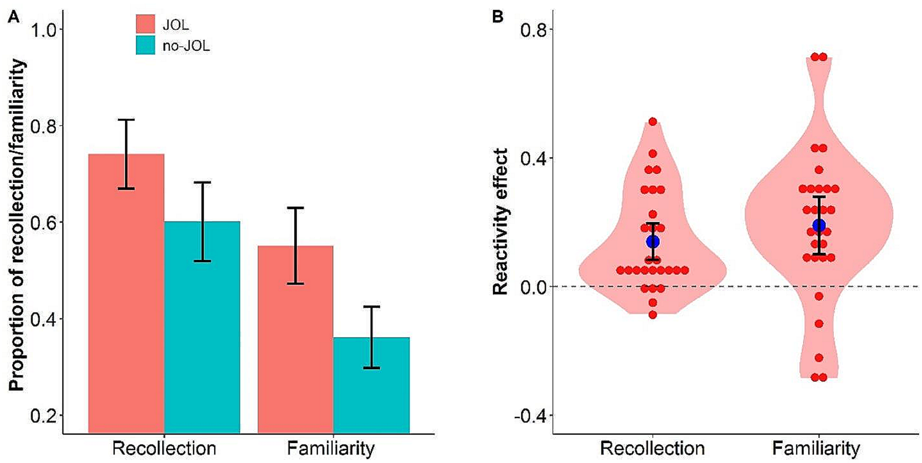

Fig. 3 Recollection and familiarity results in Experiment 2. Panel A: The proportions of recollection and familiarity for JOL and no-JOL words in Experiment 2. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in R/K response between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs

recognition memory. The proportion (89.3%; 25 out 28) of participants showing positive reactivity was substantially larger than the proportion (10.7%) showing negative reactivity, $\chi^2(1) = 17.29, p < .001$.

Recollection and familiarity responses for JOL and no-JOL words were analyzed in a 2 (study method: JOL *vs.* no-JOL) × 2 (recognition type: recollection *vs.* familiarity) repeated measures ANOVA. As shown in Fig. 3, there was a main effect of recognition type, $F(1, 27) = 22.60$, $p < .001$, $\eta_p^2 = .46$, $BF_{incl} = 426.40$, with higher proportion of recollection than that of familiarity. More importantly, there was a main effect of study method, $F(1, 27) = 30.11$, $p < .001$, $\eta_p^2 = .53$, $BF_{incl} = 1.70e+3$, with superior recognition performance in the JOL condition than in the no-JOL condition, indicating an overall positive reactivity effects of making JOLs on recollection and familiarity. Furthermore, there was no significant interaction between study method and recognition type, $F(1, 27) = 1.11$, $p = .30$, $\eta_p^2 = .04$, $BF_{incl} = 0.42$, indicating no statistically detectable difference between the reactivity effects on recollection and familiarity.

With recollection responses as the dependent variable, a paired *t*-test showed that the proportion of recollection of JOL words ($M = 0.74$, $SD = 0.19$) was significantly higher than that of no-JOL words ($M = 0.60$, $SD = 0.22$), difference = 0.14, 95% CI = [0.08, 0.20], $t(27) = 4.84$, $p < .001$, $d = 0.92$, $BF_{10} = 526.87$. The proportion (82.1%; 23 out of 28) of participants showing positive reactivity was substantially larger than the proportion (10.7%) showing negative reactivity, $\chi^2(1) = 15.39$, $p < .001$, and also substantially larger than the proportion (7.1%) showing no reactivity, $\chi^2(1) = 17.64$, $p < .001$.

With familiarity responses as the dependent variable, the proportion of familiarity of JOL words ($M = 0.55$, $SD = 0.21$) was also significantly higher than that of no-JOL words ($M = 0.36$, $SD = 0.17$), difference = 0.19, 95% CI = [0.10, 0.28], $t(27) = 4.15$, $p < .001$, $d = 0.79$, $BF_{10} = 99.97$. The proportion (82.1%; 23 out of 28) of participants showing positive reactivity was substantially larger than the proportion (17.9%) showing negative reactivity, $\chi^2(1) = 11.57$, $p < .001$.

Overall, these results demonstrate that making JOLs reactively facilitates both recollection and familiarity. Compared with not making JOLs, making JOLs significantly promoted recognition memory by improving recollection of details and familiarity of JOL words. In addition, recollection and familiarity were enhanced to a roughly equal extent.

Experiment 3

To our knowledge, Experiment 2 is the first to establish positive JOL reactivity effects on recollection and familiarity. It is necessary to test the replicability of these findings, which is the main aim of Experiment 3. The traditional R/K procedure, also known as the sequential R/K procedure (Eldridge et al., 2002; Naveh-Benjamin & Kilb, 2012), was used in Experiment 2. In the sequential R/K procedure, participants first made an old/new judgment and then made an R/K judgment (i.e., two-step). Different from Experiment 2, Experiment 3 employed another form of the R/K procedure, known as the simultaneous R/K procedure (Eldridge et al., 2002; Mulligan et al., 2010), in which participants made a single judgment (i.e., R *vs.* K *vs.* new) for each item (i.e., one-step). The use of this alternative method provided a robust extension of Experiment 2's findings.

Method

Participants

According to the effect size of the main effect of study method on remember/know responses observed in Experiment 2 ($\eta_p^2 = .53$), a power analysis showed that at least 10 participants were needed to observe a significant reactivity effect at 0.80 power. To be more conservative, we decided to increase the sample size to 25. In total, 28 participants (M_{age} = 21.32, SD=2.07; 26 female) were recruited from BNU participant pool. All participants provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, design and procedure

The materials, experimental design and procedure were identical to those in Experiment 2, but with one exception. The familiarity- and recollection-based processes were assessed using the simultaneous R/K procedure (Eldridge et al., 2002), in which participants were asked to choose one of three alternatives (Remember, Know, New). Specifically, in the final test, the 160 studied (old) and 160 new words were presented one-by-one in a random order, and participants were instructed to make a "new", "familiar", or "remember" response for each test word by pressing keycodes "J", "K", and "L". The prompts for these keycodes were always presented below the word. There was no time pressure and no feedback in the recognition test.

Participants were asked to make a "new" response when the word was believed not to have appeared in the study phase. If they recognized the word as "old", it would be classified as either "remember" or "familiar". The instructions regarding the two terms were the same as in Experiment 2. To verify compliance, participants were also asked to provide examples of each choice after the test. The results showed that all participants correctly understood and followed the task requirements.

Results and discussion

Regarding hit rates, JOL words (M=0.86, SD=0.11) were recognized more accurately than no-JOL words (M=0.80, SD=0.18), difference=0.06, 95% CI = [0.01, 0.11], $t(27)=2.66$, $p=.013$, $d=0.50$, $BF_{10}=3.66$ (see Fig. 4), replicating the positive reactivity effect on recognition memory. The proportion (67.9%; 19 out 28) of participants showing positive reactivity was substantially larger than the proportion (21.4%) showing negative reactivity, $\chi^2(1)=6.76$, $p=.009$, and also substantially larger than the proportion (10.7%) showing no reactivity, $\chi^2(1)=11.64$, $p<.001$.

Recollection and familiarity responses for JOL and no-JOL words were analyzed in a 2 (study method: JOL *vs.* no-JOL) × 2 (recognition type: recollection *vs.* familiarity) repeated measures ANOVA. As shown in Fig. 5, there was a main effect of recognition type, $F(1, 27)=27.89$, $p<.001$, $\eta_p^2=.51$, BF_{incl} = 1.31e+3, with higher proportion of recollection than that of familiarity. However, there was no main effect of study method, $F(1, 27)=1.54$, $p=.23$, $\eta_p^2=.05$, BF_{incl} = 0.44, indicating little reactive influence of JOLs on recollection and familiarity. Furthermore, there was no significant interaction between study method and recognition type, $F(1, 27)=0.20$, $p=.66$, $\eta_p^2=.01$, BF_{incl} = 0.30, indicating a minimal difference between the reactivity effects on recollection and familiarity.

With recollection responses as the dependent variable, a paired *t*-test showed that there was no detectable difference in the proportion of recollection between JOL (M=0.74, SD=0.20) and no-JOL (M=0.70, SD=0.23) conditions, difference=0.05, 95% CI = [-0.03, 0.12], $t(27)=1.22$, $p=.23$, $d=0.23$, $BF_{10}=0.39$. The proportion (57.1%; 16 out of 28) of par-

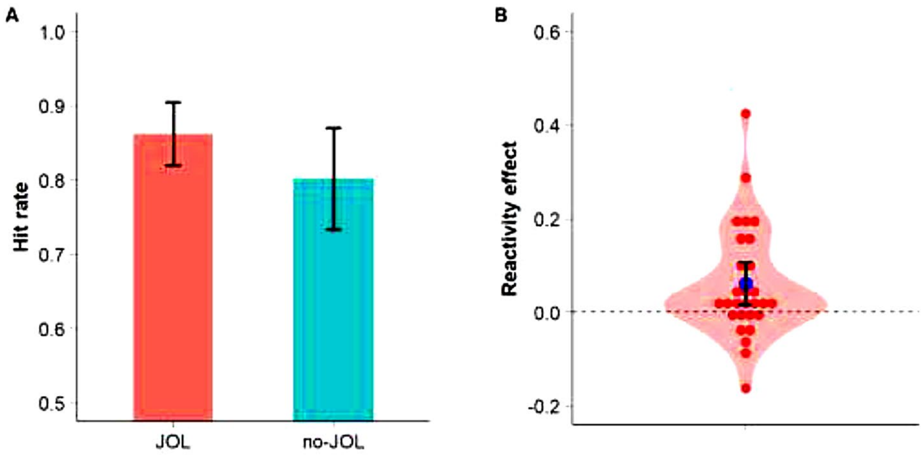

Fig. 4 Hit rate results in Experiment 3. Panel A: Hit rates for JOL and no-JOL words in Experiment 3. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in hit rates between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs

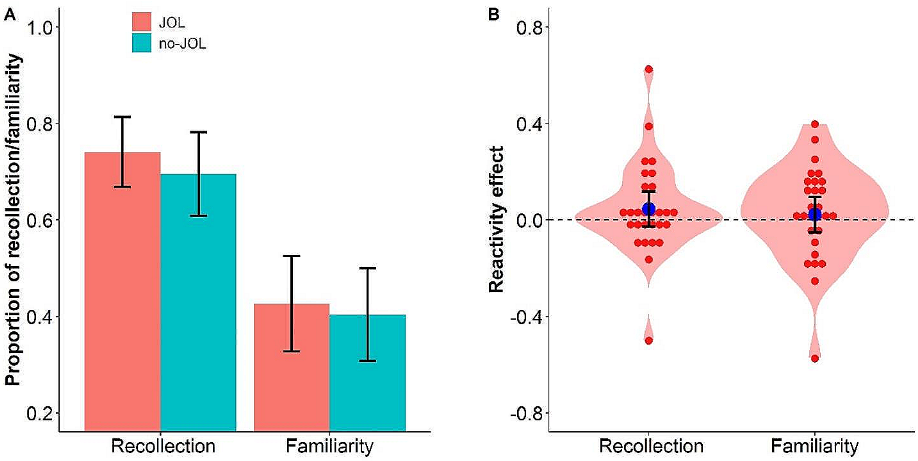

Fig. 5 Recollection and familiarity results in Experiment 3. Panel A: The proportions of recollection and familiarity for JOL and no-JOL words in Experiment 3. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in R/K response between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs.

ticipants showing positive reactivity did not significantly differ from the proportion (35.7%) showing negative reactivity, $\chi^2(1)=1.39$, $p=.24$. But both the proportions of participants showing positive, $\chi^2(1)=10.89$, $p<.001$, and negative reactivity, $\chi^2(1)=5.33$, $p=.021$, significantly larger than the proportion (7.1%) showing no reactivity.

With familiarity responses as the dependent variable, there was also no detectable difference in the proportion of familiarity between JOL ($M=0.43$, $SD=0.27$) and no-JOL

(M=0.40, SD=0.26) conditions, difference=0.02, 95% CI = [-0.05, 0.10], $t(27)=0.63$, $p=.54$, $d=0.12$, $BF_{10}=0.24$. The proportion (50.0%; 14 out of 28) of participants showing positive reactivity did not significantly differ from the proportion (32.1%) showing negative reactivity, $\chi^2(1)=1.09$, $p=.30$, but significantly larger than the proportion (17.9%) showing no reactivity, $\chi^2(1)=4.26$, $p=.039$. There was no significant difference in the proportion of participants showing negative and no reactivity, $\chi^2(1)=1.14$, $p=.29$.

Overall, Experiment 3 successfully replicated the positive reactivity effect on recognition memory. While the reactive influence of JOLs on recollection and familiarity did not attain statistical significance, the proportions of both recollection and familiarity for JOL words exhibited a consistent numerical advantage over no-JOL words. The non-significant reactivity effects on recollection and familiarity may be attributed to the low proportion of familiarity responses under the simultaneous R/K procedure. In this experiment, three participants did not make any familiarity responses, and more than half of the participants made familiarity responses for $<10\%$ of studied words. Furthermore, the simultaneous R/K procedure is posited to yield a higher proportion of recollection compared to the sequential R/K procedure (Eldridge et al., 2002), which might in turn diminish the occurrence of familiarity responses (Brown & Bodner, 2011).

Experiment 4

Previous studies have shown that when there is a long retention interval (i.e., from one day to several months) between the study and test phases, recollection responses decline sharply, while familiarity responses decline relatively slowly (Gardiner & Java, 1991; Hockley & Consoli, 1999; Meier et al., 2013). Hence, Experiment 4 was designed to decrease the proportion of recollection responses and reduce its constraining effect on familiarity responses by inserting a 24-hour interval between study and test, which allowed us to further explore the reactive influences on recollection and familiarity by using the simultaneous R/K procedure. Additionally, by prolonging the study-test interval, we aimed to examine whether the reactivity effects on recollection and familiarity are long-lasting. To our knowledge, all previous studies explored the JOL reactivity effect on short-term recognition memory in which the study-test intervals were about 2 to 10 min (Li et al., 2022; Maxwell & Huff, 2023; Myers et al., 2020; Shi et al., 2022; Zhao et al., 2022). Hence, it is critical to examine whether this enhancing effect on memory survives after a long-term delay.

Method

Participants

Consistent with Experiment 3, we decided to set the sample size to 25. Finally, 30 participants ($M_{age}=21.50$, $SD=1.85$; 22 female) recruited from the BNU participant pool. All participants provided informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, design and procedure

The materials, experimental design and procedure were identical to those in Experiment 3, but with one exception. Specifically, after the study phase, participants were dismissed and invited to come back 24 h later. One day later, they returned to the laboratory and completed the final recognition test with the simultaneous R/K procedure.

Results and discussion

Regarding hit rates, JOL words (M=0.76, SD=0.16) were recognized more accurately than no-JOL words (M=0.70, SD=0.19), difference=0.06, 95% CI = [0.02, 0.10], $t(29)=2.99$, $p=.006$, $d=0.55$, $BF_{10}=7.40$ (see Fig. 6), indicating that the positive reactivity effect on recognition memory is long-lasting (that is, it can last for at least 24 h). The proportion (66.7%; 20 out 30) of participants showing positive reactivity was substantially larger than the proportion (26.7%) showing negative reactivity, $\chi^2(1)=5.14$, $p=.023$, and also substantially larger than the proportion (6.7%) showing no reactivity, $\chi^2(1)=14.73$, $p<.001$.

Recollection and familiarity responses for JOL and no-JOL words were analyzed in a 2 (study method: JOL *vs.* no-JOL) × 2 (recognition type: recollection *vs.* familiarity) repeated measures ANOVA. As shown in Fig. 7, a main effect of recognition type was observed, $F(1, 29)=5.10$, $p=.032$, $\eta_p^2=.15$, $BF_{incl}=1.99$, with higher proportion of familiarity than that of recollection. More importantly, there was a main effect of study method, $F(1, 29)=9.34$, $p=.005$, $\eta_p^2=.24$, $BF_{incl}=8.75$, with superior recognition performance in the JOL condition than in the no-JOL condition, indicating an overall positive reactivity effects on recollection and familiarity. Notably, there was no significant interaction between study method and recognition type, $F(1, 29)=0.68$, $p=.42$, $\eta_p^2=.02$, $BF_{incl}=0.31$, indicating minimal difference between the reactivity effects on recollection and familiarity.

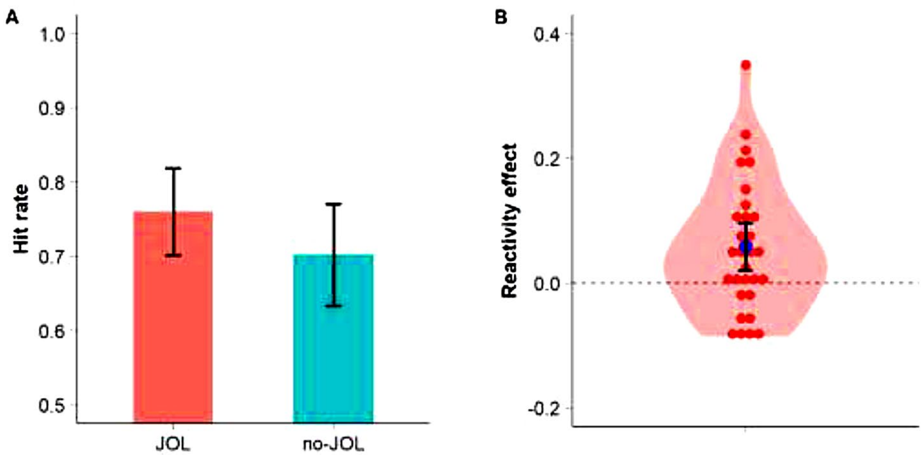

Fig. 6 Hit rates results in Experiment 4. Panel A: Hit rates for JOL and no-JOL words in Experiment 4. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in hit rates between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs

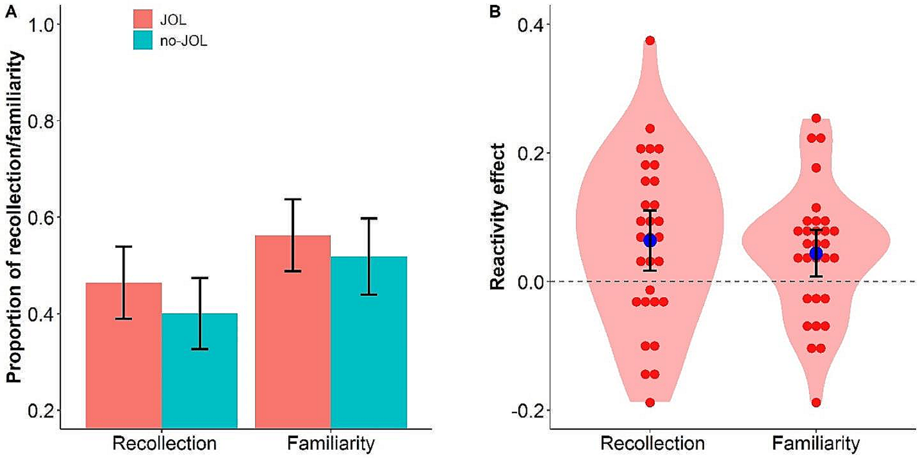

Fig. 7 Recollection and familiarity results in Experiment 4. Panel A: The proportions of recollection and familiarity for JOL and no-JOL words in Experiment 4. Panel B: Violin plot depicting the distribution of the reactivity effect of JOLs (i.e., the difference in R/K response between JOL and no-JOL words). Each red dot represents one participant's reactivity effect score and the blue point represents group average. Error bars represent 95% CIs.

With recollection responses as the dependent variable, a paired *t*-test showed that the proportion of recollection of JOL words ($M=0.46$, $SD=0.21$) was significantly higher than that of no-JOL words ($M=0.40$, $SD=0.21$), difference = 0.06, 95% CI = [0.02, 0.11], $t(29)=2.69$, $p=.012$, $d=0.49$, $BF_{10}=3.91$. The proportion (66.7%; 20 out of 30) of participants showing positive reactivity was numerically larger than the proportion (33.3%) showing negative reactivity, $\chi^2(1)=3.33$, $p=.068$.

With familiarity responses as the dependent variable, the proportion of familiarity of JOL words ($M=0.56$, $SD=0.21$) was also significantly higher than that of no-JOL words ($M=0.52$, $SD=0.22$), difference = 0.04, 95% CI = [0.01, 0.08], $t(29)=2.41$, $p=.023$, $d=0.44$, $BF_{10}=2.26$. The proportion (70%; 21 out of 30) of participants showing positive reactivity was significantly larger than the proportion (30%) showing negative reactivity, $\chi^2(1)=4.80$, $p=.028$.

Overall, the above results demonstrated that the positive reactivity effects on recollection and familiarity remained stable even in a test administered one-day later, and there was no detectable difference in the magnitude of positive reactivity effects on recollection and familiarity.

A mini meta-analysis

Experiments 2–4 consistently showed no detectable difference between the reactivity effects on recollection and familiarity. Readers may worry that the non-significant differences might result from low statistical power because the sample size in each experiment was relatively small. To mitigate potential concerns about statistical power and to make more firm conclusions, several mini random-effects meta-analyses were conducted, which integrated results across experiments. All meta-analyses were performed via the R *metafor* package (Viechtbauer, 2010).

As shown in Fig. 8, the results showed a large positive reactivity effect on recognition performance (i.e., hit rates), Hedges' $g=0.76$ [0.47, 1.05], $Z=5.09$, $p<.001$. More importantly, there was a positive reactivity effect on recollection, Hedges' $g=0.53$ [0.16, 0.90], $Z=2.83$, $p=.005$, and a positive reactivity effect on familiarity, Hedges' $g=0.39$ [0.08, 0.70], $Z=2.44$, $p=.015$.

Finally, for each participant in each of Experiments 2–4, we first calculated the difference in the proportion of recollection between the JOL and no-JOL words, which was taken as an index of the JOL reactivity effect on recollection. Then, for each participant, we calculated the difference in the proportion of familiarity between the JOL and no-JOL words, which was taken as an index of the JOL reactivity effect on familiarity. Next, for each experiment, we calculated a Cohen's *d* to measure the standardized difference between the reactivity effects on recollection and familiarity. Cohen's *d*s were then transformed to Hedges' *g*s, and submitted to a random-effects meta-analysis, which was performed to further determine whether making JOLs enhances recollection and familiarity to different extents. The results showed that there was minimal difference between reactive influence of JOLs on recollection and familiarity, Hedges' $g=0.02$ [-0.23, 0.27], $Z=0.18$, $p=.86$.

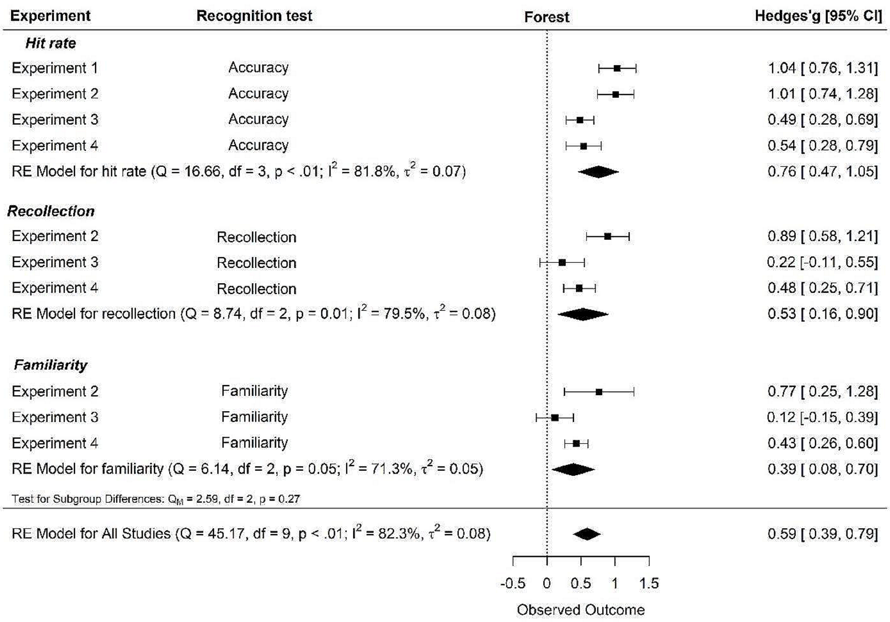

Fig. 8 Forest plot depicting the meta-analytic results. Error bars represent 95% CIs

General discussion

Recognition memory is an important part of cognitive function of humans. Previous studies have explored the JOL reactivity effect on recognition memory (Li et al., 2022; Tekin & Roediger, 2020; Zhao et al., 2022). Going beyond, the current study is the first to investigate whether making item-by-item JOLs reactively enhances recognition memory through facilitating its recollection and familiarity components. Four experiments and a meta-analysis jointly showed that making JOLs substantially improved recognition performance, recollection and familiarity, and these positive JOL reactivity effects were long-lasting (at least 24 h). Furthermore, there was minimal difference between the reactive influences on recollection and familiarity.

Similar to other encoding strategies (e.g., generation, production, elaboration) (Ozubko et al., 2012; Sheridan & Reingold, 2011, 2012), making JOLs can not only benefit recollection-based recognition but also facilitate familiarity-based recognition. The positive reactivity effect on recognition memory, as well as its stability in promoting both recollection and familiarity, suggest the potential of soliciting JOLs as an effective learning strategy for facilitating learning of simple materials such as word lists. Furthermore, the roughly equivalent reactivity effects on recollection and familiarity can explain the findings that the JOL reactivity effect was significantly larger on recognition tests than on free recall tests (Myers et al., 2020; Zhao, Li et al., 2023a). Free recall mainly relies on conscious recollection of studied items, whereas both recollection- and familiarity-based recognition contribute to recognition performance (Yonelinas, 2002). The additional boost to item familiarity induced by making JOLs may account for the inflated effect size of the JOL reactivity effect on recognition memory.

The JOL reactivity effects on recollection and familiarity are consistent with the ELE theory (Shi et al., 2022), which attributes the positive reactivity effect to the fact that making item-by-item JOLs enhances learning engagement (e.g., reducing mind wandering) and heightens learning effort. Previous evidence has shown that making JOLs can be used to maintain attention to the learning task on hand (Carpenter & Schacter, 2018; Shi et al., 2022), and reduced attention simultaneously impairs both recollection and familiarity (Curran, 2004; Gruppuso et al., 1997; Mangels et al., 2001; Parkin et al., 1995; Yonelinas, 2001). Recollection reflects the retrieval of "qualitative" information about a previous event, while familiarity reflects the assessment of "quantitative" memory strength information (Yonelinas, 2002). According to the ELE theory, participants needed to carefully analyze each study item in order to provide an appropriate JOL, which in turn increased the likelihood of recalling relevant information about JOL words. The sustained attention during learning might also contribute to the overall memory strength of JOL words, thereby aiding participants in successfully identifying studied items based on familiarity.

It is worth noting that the ELE theory does not delineate the processes or manners by which the requirement of making JOLs enhances recollection of prior contextual details about studied items. That is, what kind of processing does the increased learning engagement by making JOLs bring about? And does the act of making JOLs itself also adds something specific to studied items that facilitates participants' recollection? For instance, in post-experiment interviews of the current study, some participants reported that they remembered that they had made memory predictions for certain items, and some participants even explicitly reported that they ironically recalled some items for which they provided memory predictions. We speculate that those contextual details introduced by making JOLs may include: (1) the source memory about JOL words (i.e., information about whether a participant himself or herself had made a JOL for a given item), (2) the JOL values assigned to different items (i.e., a specific JOL value serving as an exclusive label for a certain item), and (3) the semantic information associated with the item being judged (i.e., semantically related information activated during the process of searching for mnemonic cues to provide an appropriate JOL). All these additional encoding characteristics produced by making JOL are possibly to be integrated into the item's memory representation, becoming relevant contextual details, thereby the recollection of any previous details will facilitate its recollection. Given that these factors increase the distinctiveness of JOL items compared to no-JOL items, we provisionally designate this explanation as the *distinctiveness theory*.

Another issue that needs clarification is the connection and distinction among various theories of JOL reactivity, particularly the relationship between the cue-strengthening theory and the ELE theory. Firstly, the cue-strengthening theory primarily accounts for the positive JOL reactivity effect on memory for related word pairs (Soderstrom et al., 2015) but has difficulty in explaining the positive reactivity on memory for word lists (Li et al., 2023; Zhao, Li et al., 2023a). In contrast, the ELE theory and the distinctiveness theory can jointly explain the positive reactivity effects on recollection and familiarity. Secondly, the ELE theory can be considered as the foundation, with the other two theories offering specific pathways for implementation. When learning related word pairs, enhanced learning engagement may primarily contribute to semantic relational processing between the cue and target words, facilitating cued recall (Soderstrom et al., 2015). When learning word lists, enhanced learning engagement may improve item recollection by increasing encoding distinctiveness and improve item familiarity in a direct way.

Besides the theoretical implications discussed above, the current study also brings some practical educational implications. Building upon the ELE theory of JOL reactivity, the potential benefits of making JOLs may extend beyond simple materials and laboratory environments. Research conducted in both controlled settings (Pan et al., 2020; Varao Sousa et al., 2013) and real educational contexts (Lei et al., 2018; Lindquist & McLean, 2011) has highlighted a strong correlation between learning engagement and (academic) testing performance. Indeed, mind wandering and attention lapses are common during both online and traditional classroom learning (Szpunar et al., 2013). Therefore, soliciting JOLs could act as a simple and efficient strategy to sustain learning engagement. By inserting JOL tasks into a prolonged lecture, teachers can enhance students' involvement in class activities, ultimately fostering their academic performance. According to the findings of this study, the positive JOL reactivity effects regarding recollection and familiarity may also apply to other areas such as reading. For instance, recollection contributes to explain reading comprehension (Pelegrina et al., 2023), and increasing word familiarity can help resist interference during reading (Risko et al., 2011). Overall, the educational applications of JOL reactivity warrant further exploration (Ariel et al., 2021), especially through research with greater ecological validity.

It should be acknowledged that the current study suffers from two limitations. First, our Experiments 2–4 found positive JOL reactivity effects on recognition memory and its recollection and familiarity components for young college students. It is unknown whether these positive reactivity effects can generalize to other populations. For instance, although Zhao et al. (2022) found that elementary school children exhibit a positive JOL reactivity effect on recognition memory for word lists, it remains unknown whether making JOLs promotes their recollection- and/or familiarity-based recognition. Although Tauber and Witherby (2019) found that there is no reactive influence of making JOLs on cued recall of related word pairs for older adults, the question of whether making JOLs has the potential to facilitate their recognition memory (through its recollection and/or familiarity components) has never been explored by far. Future research is encouraged to test the JOL reactivity effect on recognition memory components for young children and older adults, which may generate important implications for educators or community workers to design interventions to improve memory performance in children and older populations.

Second, although the findings documented here lay the groundwork for exploring the cognitive underpinnings by which soliciting JOLs facilitates successful recognition, empirical tests of the above theoretical explanations have not been addressed. We proposed that the ELE theory can account for why making JOLs facilitates both recollection and familiarity. Additionally, we introduced a new theoretical framework — the distinctiveness theory — to specifically describe the manner in which making JOLs enhances the likelihood of recollecting prior contextual details of studied items. However, both of these theoretical explanations have not been thoroughly examined in the current study. Shi et al. (2022) have demonstrated the role of enhanced learning engagement in the positive JOL reactivity effect on visual recognition memory for object and scene images. Future research needs to test the ELE theory's validity in explaining the reactivity effects on (familiarity-based) recognition memory for word lists, and further test the distinctiveness theory's validity in explaining the reactivity effect on recollection-base recognition.

Concluding remarks

Making item-by-item JOLs reactively facilitates both recollection and familiarity, with no detectable difference between its enhancing effects on these recognition memory components. Furthermore, these positive reactivity effects are rather durable (that is, they can last at least 24 h). The distinctiveness account and the ELE theory are potential explanations for the reactivity effects on recollection and familiarity, respectively. Further empirical tests on these theoretical accounts are called for.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11409-024-09382-1.

Acknowledgements This research was supported by the National Natural Science Foundation of China (32171045; 32371116; 32000742; 32200841), the Fundamental Research Funds for the Central Universities (1233200008), Tang Scholar Foundation, the Research Program Funds of the Collaborative Innovation Center of Assessment for Basic Education Quality at Beijing Normal University (2022-01-132-BZK01), the United Kingdom Economic and Social Research Council (ES/S014616/1), and the China Postdoctoral Science Foundation (2023M740300).

Data availability The data contained in this project are publicly available at Open Science Framework (https://osf.io/hsy6j/).

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *Neuroimage*, *62*(1), 439–450. https://doi.org/10.1016/j.neuroimage.2012.04.031.

Aggleton, J. P., Vann, S. D., Denby, C., Dix, S., Mayes, A. R., Roberts, N., & Yonelinas, A. P. (2005). Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*, *43*(12), 1810–1823. https://doi.org/10.1016/j.neuropsychologia.2005.01.019.

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, *33*(2), 693–712. https://doi.org/10.1007/s10648-020-09556-8.

Bader, R., & Mecklinger, A. (2017). Separating event-related potential effects for conceptual fluency and episodic familiarity. *Journal of Cognitive Neuroscience*, *29*(8), 1402–1414. https://doi.org/10.1162/jocn_a_01131.

Bastin, C., Besson, G., Simon, J., Delhaye, E., Geurten, M., Willems, S., & Salmon, E. (2019). An integrative memory model of recollection and familiarity to understand memory deficits. *Behavioral and Brain Sciences*, *42*, https://doi.org/10.1017/S0140525X19000621. e281, Article e281.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*(6), 897–903. https://doi.org/10.3758/s13421-013-0307-8.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823.

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*(1), 51–61. https://doi.org/10.1038/35049064.

Brown, A. A., & Bodner, G. E. (2011). Re-examining dissociations between remembering and knowing: Binary judgments vs. independent ratings. *Journal of Memory and Language*, *65*(2), 98–108. https://doi.org/10.1016/j.jml.2011.04.003.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS One*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729.

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*(7), 988–1004. https://doi.org/10.1037/xge0000391.

Cohen, M. S., Rissman, J., Hovhannisyan, M., Castel, A. D., & Knowlton, B. J. (2017). Free recall test experience potentiates strategy-driven effects of value on memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(10), 1581–1601. https://doi.org/10.1037/xlm0000395.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia*, *42*(8), 1088–1106. https://doi.org/10.1016/j.neuropsychologia.2003.12.011.

Curran, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. *Neuroimage*, *36*(2), 464–471. https://doi.org/10.1016/j.neuroimage.2006.12.016.

Dobbins, I. G., Kroll, N. E. A., & Yonelinas, A. P. (2004). Dissociating familiarity from recollection using rote rehearsal. *Memory & Cognition*, *32*(6), 932–944. https://doi.org/10.3758/BF03196871.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory (Hove, England)*, *26*(6), 741–750. https://doi.org/10.1080/09658211.2017.1404111.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328.

Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, *9*(1), 139–145. https://doi.org/10.3758/BF03196270.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology*, *70*(2), 99–115. https://doi.org/10.1037/cep0000089.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, *19*(6), 617–623. https://doi.org/10.3758/BF03197157.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(2), 259–278. https://doi.org/10.1037/0278-7393.23.2.259.

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition*, *27*(4), 657–664. https://doi.org/10.3758/BF03211559.

Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PloS One*, *10*(11), e0142351–e0142351. https://doi.org/10.1371/journal.pone.0142351.

Jacoby, L., Yonelinas, A., & Jennings, J. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen, & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Erlbaum.

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*, *25*(6), 2356–2364. https://doi.org/10.3758/s13423-018-1463-4.

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

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality*, *46*(3), 517–528. https://doi.org/10.2224/sbp.7054.

Li, B., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., Yin, Y., Liu, M., Yang, C., & Luo, L. (2022). Soliciting judgments of forgetting reactively enhances memory as well as making judgments of learning: Empirical and meta-analytic tests. *Memory & Cognition*, *50*(5), 1061–1077. https://doi.org/10.3758/s13421-021-01258-y.

Li, B., Zhao, W., Shi, A., Zhong, Y., Hu, X., Liu, M., Luo, L., & Yang, C. (2023). Does the reactivity effect of judgments of learning transfer to learning of new information? *Memory*, 1–13. https://doi.org/10.1080/09658211.2023.2208792.

Lindquist, S. I., & McLean, J. P. (2011). Daydreaming and its correlates in an educational environment. *Learning and Individual Differences*, *21*(2), 158–167. https://doi.org/10.1016/j.lindif.2010.12.006.

Mangels, J. A., Picton, T. W., & Craik, F. I. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research*, *11*(1), 77–95. https://doi.org/10.1016/S0926-6410(00)00066-5.

Maxwell, N. P., & Huff, M. J. (2023). Judgment of learning reactivity reflects enhanced relational encoding on cued-recall but not recognition tests. *Metacognition and Learning*. https://doi.org/10.1007/s11409-023-09369-4.

Meier, B., Rey-Mermet, A., Rothen, N., & Graf, P. (2013). Recognition memory across the lifespan: The impact of word frequency and study-test interval on estimates of familiarity and recollection. *Frontiers in Psychology*, *4*, 787–787. https://doi.org/10.3389/fpsyg.2013.00787.

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*(2), 200–219. https://doi.org/10.1037/a0039923.

Mulligan, N. W., Besken, M., & Peterson, D. (2010). Remember-know and source memory instructions can qualitatively change old-new recognition accuracy: The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(2), 558–566. https://doi.org/10.1037/a0018408.

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*(5), 745–758. https://doi.org/10.3758/s13421-020-01025-5.

Naveh-Benjamin, M., & Kilb, A. (2012). How the measurement of memory processes can affect memory performance: The case of remember/know judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(1), 194–203. https://doi.org/10.1037/a0025256.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science*, *2*(4), 267–270. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In H. B. Gordon (Ed.), *Psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1.

Pan, S. C., Sana, F., Schmitt, A. G., & Bjork, E. L. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition*, *9*(4), 542–554. https://doi.org/10.1016/j.jarmac.2020.07.004.

Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness and Cognition*, *4*(4), 387–398. https://doi.org/10.1006/ccog.1995.1046.

Pelegrina, S., Mirandola, C., Linares, R., & Carretti, B. (2023). Recollection contributes to children's reading comprehension: Using the process dissociation procedure in a working memory updating task. *Journal of Experimental Child Psychology*, *226*, 105550. https://doi.org/10.1016/j.jecp.2022.105550.

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

Risko, E. F., Stolz, J. A., & Besner, D. (2011, Jan 20). Basic processes in reading: On the relation between spatial attention and familiarity. *Language and Cognitive Processes, 26*(1), 47–62. https://doi.org/10.1080/01690961003679574.

Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory (Hove, England)*, *29*(10), 1342–1353. https://doi.org/10.1080/09658211.2021.1985143.

Rosenstreich, E., & Ruderman, L. (2017). A dual-process perspective on mindfulness, memory, and consciousness. *Mindfulness*, *8*(2), 505–516. https://doi.org/10.1007/s12671-016-0627-4.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7.

Sheridan, H., & Reingold, E. M. (2011). Recognition memory performance as a function of reported subjective awareness. *Consciousness and Cognition*, *20*(4), 1363–1375. https://doi.org/10.1016/j.concog.2011.05.001.

Sheridan, H., & Reingold, E. M. (2012). Levels of processing influences both recollection and familiarity: Evidence from a modified remember-know paradigm. *Consciousness and Cognition*, *21*(1), 438–443. https://doi.org/10.1016/j.concog.2011.09.022.

Shi, A., Xu, C., Zhao, W., Shanks, D. R., Hu, X., Luo, L., & Yang, C. (2022). Judgments of learning reactively facilitate visual memory by enhancing learning engagement. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-022-02174-1.

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia*, *45*(10), 2163–2179. https://doi.org/10.1016/j.neuropsychologia.2007.03.007.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.org/10.1037//0278-7393.4.6.592.

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*(2), 553–558. https://doi.org/10.1037/a0038388.

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*(5), 315–316. https://doi.org/10.1111/j.1467-9280.1992.tb00680.x.

Staresina, B. P., Fell, J., Do Lam, A. T. A., Axmacher, N., & Henson, R. N. (2012). Memory signals are temporally dissociated in and across human hippocampus and perirhinal cortex. *Nature Neuroscience*, *15*(8), 1167–1173. https://doi.org/10.1038/nn.3154.

Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: From the classroom to online learning. *Frontiers in Psychology*, *4*, 495–495. https://doi.org/10.3389/fpsyg.2013.00495.

Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, *34*(6), 836–847. https://doi.org/10.1037/pag0000376.

Tekin, E., & Roediger, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift für Psychologie*, *228*(4), 278–290. https://doi.org/10.1027/2151-2604/a000425.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017.

Varao Sousa, T. L., Carriere, J. S. A., & Smilek, D. (2013). The way we encounter reading material influences how frequently we mind wander. *Frontiers in Psychology*, *4*, 892–892. https://doi.org/10.3389/fpsyg.2013.00892.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03.

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, *1122*(1), 161–170. https://doi.org/10.1016/j.brainres.2006.09.023.

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*(4), 496–503. https://doi.org/10.1016/j.jarmac.2017.08.004.

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, *1100*(1), 125–135. https://doi.org/10.1016/j.brainres.2006.05.019.

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

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 cs of recognition memory. *Journal of Experimental Psychology: General*, *130*(3), 361–379. https://doi.org/10.1037/0096-3445.130.3.361.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*, *34*(5), 622–643. https://doi.org/10.1006/jmla.1995.1028.

Zhao, W., Li, B., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., Yin, Y., Luo, L., & Yang, C. (2022). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children's learning. *Child Development*, *93*(2), 405–417. https://doi.org/10.1111/cdev.13689.

Zhao, W., Li, J., Shanks, D. R., Li, B., Hu, X., Yang, C., & Luo, L. (2023a). Metamemory judgments have dissociable reactivity effects on item and interitem relational memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *49*(4), 557–574. https://doi.org/10.1037/xlm0001160

Zhao, W., Yin, Y., Hu, X., Shanks, D. R., Yang, C., & Luo, L. (2023b). Memory for inter-item relations is reactively disrupted by metamemory judgments. *Metacognition and Learning*, *18*(2), 549–566. https://doi.org/10.1007/s11409-023-09340-3.

Authors and Affiliations

Jun Zheng[1,2] · Baike Li[3] · Wenbo Zhao[4] · Ningxin Su[2] · Tian Fan[2] · Yue Yin[2] · Yali Hu[2] · Xiao Hu[5] · Chunliang Yang[5,6] · Liang Luo[6,7]

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

✉ Liang Luo
luoliang@bnu.edu.cn

1 Center for Teacher Education Research, Faculty of Education, Beijing Normal University, Beijing 100875, China

2 Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing 100875, China

3 School of Psychology, Liaoning Normal University, Dalian 116029, China

4 School of Social Development and Public Policy, Beijing Normal University, Beijing 100875, China

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

6 Institute of Developmental Psychology, Faculty of Psychology, Beijing Normal University, 19 Xinjiekou Wai Street, Beijing 100875, China

7 State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, 19 Xinjiekou Wai Street, Beijing 100875, China