Metacognition and Learning (2025) 20:32
https://doi.org/10.1007/s11409-025-09439-9

Soliciting judgments of learning reactively enhances accessibility but undermines precision of continuous color memory

Huafeng Tang[1] · Wenbo Zhao[2] · Baike Li[3] · Shaohang Liu[1] · Anran Li[1] · Mengqi Hu[1] · Xiao Hu[4] · Liang Luo[1,5] · Chunliang Yang[1,4]

Received: 7 May 2024 / Accepted: 29 July 2025

Abstract
An emerging body of studies has observed that soliciting judgments of learning (JOLs) reactively changes recall or recognition of specific study items, a phenomenon known as the *reactivity effect* of JOLs on memory. The current studies explored whether soliciting JOLs reactively affects continuous color memory and whether it affects memory accessibility or memory precision (or a combination of both). Experiment 1 employed a classical continuous color memory task in which participants studied animal images in different colors and then reconstructed the colors on a continuous matching wheel, and found that making JOLs reactively enhanced accessibility but impaired precision of memory. Experiment 2 replicated these dissociated reactivity effects and further found that articulatory suppression successfully eliminated these effects, suggesting that making JOLs reactively alters color memory through prompting individuals to favor the verbal-labeling strategy for encoding and retrieving visual information. The findings support the strategy-change theory to account for the JOL reactivity effect.

Keywords Judgments of learning · Reactivity effect · Continuous color memory · Accessibility and precision · Verbal-labeling · Strategy-change

Author Note: The data contained in this project are publicly available at Open Science Framework (https://osf.io/m7df6/).

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

1 Institute of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China

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

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

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

5 State Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Published online: 29 August 2025

Metacognition is an advanced ability concerning self-awareness, which plays a crucial role in the processes of learning and memory (for reviews, see Rhodes, 2016; Yang et al., 2021). Previous research has frequently employed judgments of learning (JOLs; prospective estimates of the likelihood of remembering studied information in a future occasion) as a measurement tool to assess people's metamemory monitoring ability (Murphy et al., 2022; Myers et al., 2020; Rhodes & Castel, 2009; Rhodes, 2016; Tauber & Rhodes, 2012; Yang et al., 2021). However, an emerging body of studies has revealed that making JOLs can reactively alter memory itself, a phenomenon termed as the *reactivity effect* (Double & Birney, 2018a; Li et al., 2021; Rivers et al., 2021; Senkova & Otani, 2021; Zhao et al., 2023a).

As an illustration, Soderstrom et al. (2015, Experiment 1 A) instructed two groups (JOL vs. no-JOL) of participants to study a mixed list of strongly (e.g., *pledge-promise*) and weakly (e.g., *mercy-justice*) related word pairs. In the JOL group, participants were instructed to make a JOL while studying each word pair, whereas those in the no-JOL group were not required to make concurrent JOLs. All participants then received a cued-recall test (e.g., *pledge-?*). On that test, the JOL group recalled significantly more strongly related pairs and numerically more weakly related pairs than the no-JOL group, reflecting a positive reactivity effect on memory (for related findings, see Janes et al., 2018; Li et al., 2021; Mitchum et al., 2016; Myers et al., 2020; Rivers et al., 2021; Soderstrom et al., 2015; Spellman & Bjork, 1992).

Recent studies showed that some factors (such as material type, participant population, and test format) can moderate the JOL reactivity effect. For instance, previous studies found that making JOLs reactively enhances recall or recognition of related word pairs (e.g., Li et al., 2021; Maxwell & Huff, 2022; Rivers et al., 2021), identical word pairs (Halamish & Undorf, 2023), word lists (Li et al., 2021; Zhao et al., 2022, 2023b) and visual images (Shi et al., 2023). However, making JOLs produces minimal reactive influence on recall of text passages (Ariel et al., 2021; Zhao et al., 2023a) and general knowledge facts (Schäfer & Undorf, 2023). The reactivity effect generalizes to elementary school children (Zhao et al., 2022) and young adults (Witherby & Tauber, 2017), but tends not to generalize to older adults (Tauber & Witherby, 2019). Furthermore, Myers et al. (2020) found that the reactivity effect exists in recognition and cued recall tests, but not in free recall test. Specifically, in cued recall tests, Mitchum et al. (2016) observed a positive reactivity effect for related word pairs and a negative reactivity effect for unrelated pairs.

Previous studies have found JOL reactivity effects using a variety of test formats including old/new recognition (Li et al., 2021), forced-choice recognition (Shi et al., 2023), cued recall (Koriat et al., 2004). It should be highlighted that previous reactivity studies have generally measured memory performance in a "all-or-none" manner. However, a binary measure fails to fully capture the fine-grained variations of memory recollection. Increasing evidence suggests that memory, particularly visual memory, is not a dichotomy of "all-or-none" (Berens et al., 2020; Harlow & Donaldson, 2013; Van den Berg et al., 2012). Our visual memory can retain a vast number of objects with vivid details (Brady et al., 2008; Hollingworth, 2004; Konkle et al., 2010).

To better capture the multifaceted and graded nature of memory, researchers developed continuous measures of visual memory, reflecting that an item can be partially remembered (Bays et al., 2009; Zhang & Luck, 2008). Specifically, participants are typically asked to study a set of object images presented in different colors and then reconstruct the colors of studied objects on a continuous matching wheel. Previous studies demonstrated that memory accessibility (i.e., the probability of successful retrieval) and memory precision (i.e., the fidelity of stored information) are separable components of memory (Harlow & Donaldson, 2013; Korkki et al., 2020; Onyper et al., 2010; Yonelinas & Parks, 2007), which can be selectively affected by different experimental manipulations (Liu et al., 2023; Sutterer & Awh, 2016), brain stimulation (Nilakantan et al., 2017), and developmental condition (Cooper et al., 2017). Essentially, memory accessibility is about how accessible the studied information is on a later test, whereas memory precision represents the degree of accuracy of retrieved information by comparison with the original information.

Similarly, a recent study has explored the reactivity effects on two distinct subcomponents of memory: item-specific processing and inter-item relational processing (Zhao et al., 2023b). Specifically, in Zhao et al. (2023b) Experiment 1, participants studied word lists consisting of semantically related category exemplars, which were presented in a category-blocked order to increase inter-target relations. The result showed a negative reactivity effect in memory for inter-item relations. However, in their Experiment 3, where category exemplars were presented in a random sequence to minimize semantic clustering, making JOLs improved recognition of target words, reflecting a positive reactivity effect in item memory. Zhao et al. (2023b) concluded that the dissociated reactivity effects in different memory components come from a shift in cognitive resources between item-specific processing and inter-item relational processing.

Different from Zhao et al. (2023b) explanation of memory separation, the dual-trace theory hypothesizes that visual memory draws from both visual representations and verbal labels, which jointly consume cognitive resources and may compete with one another (Donkin et al., 2015). While memory precision is mainly driven by the fidelity of visual representations, memory accessibility largely relies on the strength of verbal labels (see below for a detailed discussion; Fougnie et al., 2016; Hardman & Cowan, 2015; Palmer et al., 2015; Swan et al., 2016). Given the divergent nature of these two subcomponents of visual memory, it is possible that making JOLs may produce differential reactive influences on them.

Reactivity effects on accessibility and precision in color memory

Many studies have established that soliciting JOLs can reactively alter memory performance when using a binary measure. Different from previous studies, the current research aimed to employ a continuous color memory task, adapted from recent work in long-term continuous color memory (Sutterer & Awh, 2016), to explore whether soliciting JOLs reactively changes memory accessibility or memory precision (or a combination of both). To achieve this aim, participants in the current study were asked to study a set of object images presented in different colors in preparation for a color memory test. An overview of the procedure is shown in Fig. 1. For half of the images, they needed to make concurrent JOLs, but not for the other half. Then, they completed a color reproduction test, in which they needed to recall the color of a studied object by selecting a color on a 360-degree color wheel as shown in Fig. 1 (Brady et al., 2013; Korkki et al., 2020; Sutterer & Awh, 2016).

Response error on each trial was quantified by computing the angular disparity between the studied value (i.e., the studied color) and the response value (i.e., the selected color in the color reproduction test). The magnitude of recall error spanned from 0°, representing

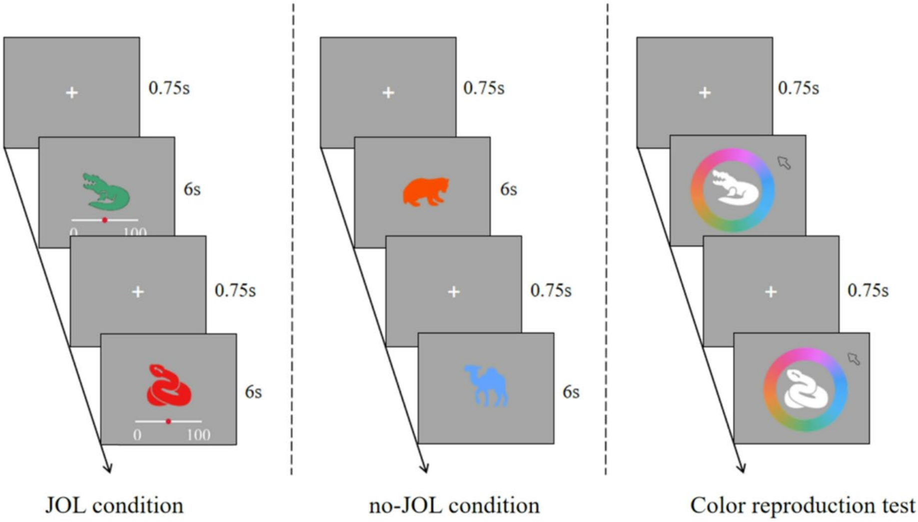

Fig. 1 Flow chart depicting the stimuli and task procedure used in Experiment 1

a flawless response, to ±180°, denoting a maximally imprecise response. On some trials, participants might fail to recall the color associated with the shape cue, leading them to make random guesses regarding the target color. Random guesses would result in a uniform distribution of response errors (Korkki et al., 2020; Sutterer & Awh, 2016; Zhang & Luck, 2008). On other trials, participants might successfully remember the studied color, their color responses should be centered around the correct color value with some degree of error. Put differently, if the color of a studied image has been successfully remembered, the recalled color should be close to (or even exactly same as) the original color. In such a situation, the distribution of response errors can be effectively characterized by a von Mises distribution, which serves as the circular analogue of a Gaussian distribution, given the circular nature of the tested color space.

These two types of trials (i.e., remembered and forgotten items) are mixed together (Zhang & Luck, 2008; see Fig. 2a), and hence the distribution of response errors can be well described by a mixture of uniform and von Mises distributions that reflects accessibility (P_{mem}) and precision (SD_{mem}) of continuous color memory (Zhang & Luck, 2008; see Fig. 2b). P_{mem} is calculated as the inverse of the height of the uniform distribution, which is operationalized as 1 − the proportion of random guesses, with larger P_{mem} values representing greater levels of memory accessibility (i.e., low proportion of random guesses in the color reproduction test). SD_{mem} denotes the standard deviation of the von Mises distribution, with smaller SD_{mem} values representing greater levels of memory precision (i.e., the distribution of response errors was more centered and less dispersed). These are two main measures were used to test which components of continuous color memory are reactively altered by making JOLs. Previous work showed that making JOLs reactively facilitates the probability of successfully remembering visual information by enhancing learning engagement (Shi et al., 2023). As mentioned before, the shift of cognitive resources may impair the fidelity of the stored information. Hence, it could be expected that making JOLs might reactively enhance P_{mem} but undermine SD_{mem}.

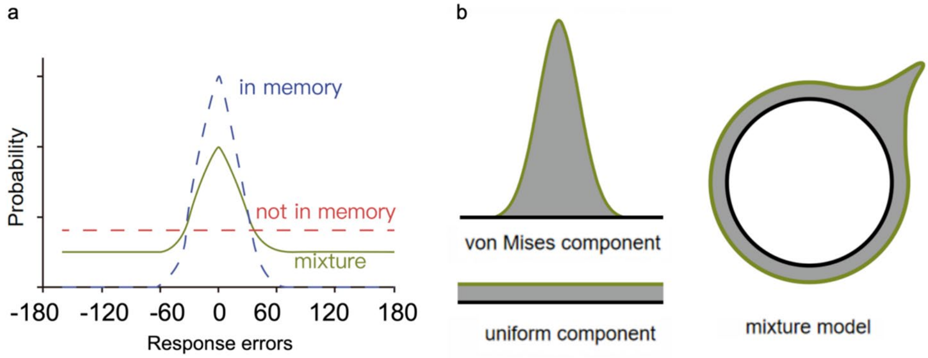

Fig. 2 Illustration of the probabilistic mixture model. *Note*. (**a**) Illustration of distribution of response errors in the color reproduction test. If the color of the probed item is retained in memory, the reported color usually approximates the original color, and response errors should follow a von Mises distribution (blue line). Conversely, if the color of the probed item is forgotten, any color value is equally likely to be selected, and response errors should follow a uniform distribution (red line). When aggregating across trials, the data represent a mixture of these two trial types (green line), adjusted by the probability of the color of the probed item being stored in memory. (**b**) The probabilistic mixture model comprises a combination of a von Mises distribution centered around the target feature value and a circular uniform distribution. Memory accessibility is defined as the inverse of the height of the uniform distribution, and memory precision is quantified as the concentration of the von Mises distribution

Strategy change as an explanation for reactivity effects

Besides exploring the JOL reactivity effect on continuous color memory, the current research also targeted to test a recently proposed theory of the JOL reactivity effect, that is, the *strategy-change theory* (Mitchum et al., 2016; Rivers et al., 2021; Sahakyan et al., 2004; Shi et al., 2023). This theory hypothesizes that asking participants to make JOLs enhances participants' awareness of task difficulty, the effectiveness of employed study strategies, the gap between current level of knowledge mastery and their desired learning goals, and so on. Enhanced metacognitive awareness may then prompt them to change their study strategies (e.g., changing from less effective ones to those more effective) to complete the learning task, which then alters their ultimate learning performance and induces a reactivity effect. Suggestive evidence supporting this hypothesis comes from Sahakyan et al. (2004), who found that asking participants to make a JOL (i.e., predicting the number of words they would remember in a later memory test) after studying a word list caused them to abandon less effective strategies, such as rote rehearsal, and utilize more effective strategies (based on participants' retrospective reports) to encode a subsequent list of words. In contrast, when making JOLs was not required, participants continued relying on ineffective strategies (e.g., rote rehearsal) to study the subsequent list. Based on participants' retrospective reports of strategy changes from a prior work (Sahakyan & Delaney, 2003), Sahakyan et al., (2004) proposed that making JOLs may reactively enhance memory by inducing a change in study strategies across tasks, although they did not directly assess participants' strategy use.

Subsequent studies have provided inconsistent support for the strategy-change theory. Two studies failed to see differences in self-reported strategy use across JOL and no JOL conditions, although they did see positive reactivity effects in the JOL condition (Mitchum et al., 2016; Rivers et al., 2021). For instance, Mitchum et al. (2016) asked two groups of participants to study a mixed list of related and unrelated word pairs. Half of the participants were required to make JOLs during studying, while the other half studied all word pairs without making JOLs. At the very end of the experiment, participants were asked to fill out a learning strategy questionnaire to report the study strategies (e.g., sentence-making, mental imaging, and mechanical repetition) they used in the JOL and no-JOL conditions. They were then asked to report the frequency of each strategy they used on a scale ranging from 0 (*never used*) to 10 (*used for studying every item*). Mitchum et al. (2016) observed a positive reactivity effect for related word pairs and a negative reactivity effect for unrelated pairs. Of critical interest, there was no difference in the frequency of study strategy usage between the JOL and no-JOL conditions. Similar findings were documented by Rivers et al. (2021), which examined the differences in study strategy usage between the JOL and no-JOL conditions at the item level. In this study, participants studied related and unrelated word pairs and made JOLs for a randomly selected half of the word pairs but not for the other half. During the final test and after answering each test question, participants were asked to report which study strategies they used to study the just-tested item. The results showed that making JOLs reactively enhances recall of related word pairs. Consistent with Mitchum et al. (2016), Rivers et al. (2021) found no difference in study strategy usage between the JOL and no-JOL conditions, again failing to provide evidence for the strategy-change theory.

In contrast, Shi et al. (2023) provided evidence supporting the strategy-change theory. Specifically, Shi et al. (2023) instructed all participants to learn four lists of visual images, and participants made JOLs for two lists of images but not for the other two lists. Following the study phase, participants undertook a recognition test on all studied images. After the recognition test, participants who exhibited positive reactivity (66.7%) were asked to explain why making JOLs enhanced their memory, while those who showed negative reactivity (26.2%) explained why making JOLs impaired their memory. Participants who showed no reactivity (7.1%) explained why making JOLs had no impact on their memory. The results showed that JOL images were recognized more accurately than no-JOL ones, reflecting a positive reactivity effect on visual memory. More importantly, among the participants who showed a positive reactivity effect, 41.1% reported that making JOLs improved their memory performance through prompting them to use better study strategies (e.g., searching for distinctive features, focusing more on visual details, and self-evaluation).

It should be highlighted that all three studies, despite reaching inconsistent conclusions, used retrospective self-reports to test the strategy-change theory. The nature of the retrospective strategy questions differed across three studies. Mitchum et al. (2016) asked participants to rate the frequency of a list of strategies in each of the JOL and no-JOL conditions. Rivers et al. (2021) collected participants' open-ended strategy reports for each item. Shi et al. (2023) asked a single, direct question about why making JOLs affected their memory, without prompting comparisons across items or conditions. It is well-known that self-reports suffer from a variety of biases and illusions (Nelson, 1990; Senkova & Otani, 2021; Yang et al., 2018), and may not accurately reflect the reality. Different from previous studies, the current studies used a more objective measure, that is, the level of divergence between the response distribution and a uniform distribution (Parameter D; see below for details), to test the strategy change theory of JOL reactivity.

Experiment 1

Experiment 1 was conducted to explore (1) whether making JOLs reactively changes continuous color memory, (2) whether it changes memory performance through affecting memory accessibility or precision (or a combination of both), and (3) whether it changes participants' study strategies.

Methods

Participants

A pilot study with 10 participants was conducted to ascertain the required sample size. We conducted comparative analyses of participants' color reproduction responses in the JOL and no-JOL conditions using mixture model at both the aggregated and individual levels (see below for details). Since individual-level analyses require *t*-tests, the power analysis was based on these *t*-test results of the pilot study. The pilot results of individual-level analyses revealed a moderate effect size for the reactivity effect on P_{mem} (Cohen's $d=0.594$) and a large effect size for the effect on SD_{mem} (Cohen's d = -0.908). Detailed results of the pilot study are reported in the Appendix. A power analysis, conducted via G*power (Faul et al., 2007), indicated that 25 participants were required to detect a significant (two-tailed $\alpha=0.05$) reactivity effect on P_{mem} with a power of 0.80.

Accordingly, 25 participants were recruited from the participant pool at Beijing Normal University. Two extra participants were recruited as replacements for participants whose data were excluded because their estimated parameters of P_{mem} or SD_{mem} deviated by more than three standard deviations (*SD*s) from the group mean (Korkki et al., 2020), and the removal criteria were set before data collection. In total, the final data came from 25 participants (M_{age} = 22.20, $SD=2.14$; 17 female). All participants provided informed consent, were tested individually in a sound-proofed cubicle, reported normal or corrected-to-normal vision, and received 50 RMB as task compensation.

The current study received approval from the Ethics Committee of Faculty of Psychology at Beijing Normal University.

Materials

Two hundred and four animal images were obtained through a royalty-free clip art search on the Internet, and their original colors were removed, leaving only the shape contours (see Fig. 1). All images have been made publicly available via Open Science Framework (OSF: https://osf.io/m7df6/). Among these images, four were used for practice, with the other 200 used in the main experiment. Additionally, we established a continuous color wheel by using the Hue, Saturation, Value (HSV) color model (Smith, 1978), and selected 200 colors at an interval of 1.8° on the wheel. For each participant, each of these 200 colors was randomly assigned to a given animal image. Furthermore, for each participant, the 200 images (i.e., 200 images of colored animal shapes) were randomly divided to four lists, with 50 images in each list. Two lists were randomly assigned to the JOL condition, with the remaining two to the no-JOL condition. The presentation order of the four lists and the order of images in each list were also randomized. All stimuli were presented via *PsychoPy-2022.2.4* (Peirce, 2007).

Design and procedure

Experiment 1 involved a within-subjects (study method: JOL vs. no-JOL) design. Participants were informed that they would study four lists of colored animal images in preparation for a later color reproduction test, in which they needed to click the color on the matching wheel that they perceived as most similar to the study color for each animal shape. Participants were further informed that they needed to make item-by-item memory predictions (i.e., estimating the likelihood of remembering the studied color of each image in a later test) for two randomly chosen lists of images. For the other two lists of images, they did not need to make memory predictions. They were also explicitly instructed to memorize all images equally well irrespective of whether they need to make memory predictions or not, because all 200 images would be tested eventually.

Before commencing the formal experiment, participants completed a practice task to familiarize themselves with the experimental procedure, in which they studied and were tested on two images under each of the JOL and no-JOL conditions. Then, the main experiment began. The task procedure is depicted in Fig. 1.

At the beginning of each image list, the computer informed participants whether they needed to make memory predictions for that specific list. In a no-JOL list, the 50 images were presented one-by-one in a random order. Before presenting each image, a cross sign was shown at the center of the screen for 0.75 s to mark the inter-stimulus interval (ISI). Then the image appeared at the center of the screen for 6 s for participants to study. Participants were asked to memorize the corresponding color of the presented animal shape. After 6 s, the image automatically disappeared. Then, the next study trial started. This cycle repeated until participants had studied all images in that list.

The procedure in the JOL condition was the same as that in no-JOL condition, but with one difference. That is, during the study phase of each trial, a slider, ranging from 0 (*Sure I will not remember*) to 100 (*Sure I will remember*), was presented below the image. Participants were asked to predict how likely they would remember the on-screen shape's color in a later color reproduction test. They made their JOLs by dragging and clicking the slider scale. If they successfully made a JOL during the 6 s time window, the image remained on screen for the left duration of 6 s to ensure that the total exposure time for each trial was equal between the JOL and no-JOL conditions. If they failed to make a JOL during 6 s, a message box would appear to remind them to carefully make predictions during the required time window for subsequent images.[1] Participants needed to click the mouse to remove the message box and trigger the next trial.

After studying all four lists, participants engaged in a distractor task, in which they needed to solve as many mathematics problems (e.g., $7+45=$ __) as possible within 5 min. Then, all participants undertook a color reproduction test, in which the 200 studied animal shapes were presented one-by-one in a random order. For each test trial, a shape contour (without color) was displayed at the center of the screen, encircled by a color wheel (see Fig. 1). Participants were instructed to select the color they remembered for the on-screen shape. By moving the mouse cursor around the color wheel, the animal shape would be filled with the color at the current position of the mouse cursor. When participants were sat-

[1] We restricted the time for making JOLs at 6 s to control the total task duration between the two conditions. According to other studies conducted in our laboratory, most participants could successfully make a JOL within 2 s even when there is no time pressure.

isfied with the selected color, they pressed the left mouse button to confirm their response. After a delay of 0.5 s, the next trial commenced automatically. This cycle repeated until participants completed all test trials. There were no time pressure and no feedback in the color reproduction test.

It should be noted that the measurement scales of JOLs and memory performance were different. Specifically, JOLs were measured on a 0-100 scale, whereas memory for studied colors was measured on a 360-dgree color wheel.

To put JOLs and memory performance in the same scale, we converted memory performance to a 0-100 scale based on the number of intervening colors, rather than calculating them in degrees. This simplifies interpretation: a memory performance score of 0 reflects that the recalled color completely matched the original (studied) color, and each adjacent color represents an error score of 1 (instead of 1.8 degrees). This 0-100 memory performance scale aligns well with the 0-100 JOL scale, ensuring that differences in measurement scales do not affect readers' interpretation of our findings.

Data analysis method

Mixture-modeling analysis allowed us to determine the probability of successful retrieval (i.e., memory accessibility, measured as P_{mem}) and the fidelity (i.e., memory precision, measured as SD_{mem}) of stored information in each of the JOL and no-JOL conditions. The distribution of response errors across all participants were subjected to a Markov Chain Monte Carlo (MCMC) analysis, implemented by the "memfit" function of MATLAB *Memtoolbox* (Suchow et al., 2013), a specialized tool for analyzing mixture modeling data. The MCMC method repeatedly samples parameter values based on their ability to effectively explain the data while considering the prior (Suchow et al., 2013). Through this process, Maximum A Posteriori (MAP) estimates of two parameters (i.e., P_{mem} and SD_{mem}) and their 95% credibility intervals (CrIs) were obtained. Parameters with overlapping CrIs are referred to as "not statistically different", and parameters with non-overlapping CrIs are referred to as "statistically different". Unlike confidence intervals (CIs), Bayesian CrIs are not always symmetrical (Liu et al., 2023; Sutterer & Awh, 2016).

Aside from aggregated mixture-modeling analysis, we also performed individual-level mixture-modeling analysis to estimate P_{mem} and SD_{mem} for each participant in each of the JOL and no-JOL conditions. Frequentists and Bayesian paired *t*-tests were conducted via JASP 0.19.3.0 to compare P_{mem} and SD_{mem} between the JOL and no-JOL conditions. Individual-level modeling-analysis was performed to test the robustness of the results generated from aggregated mixture-modeling analysis.

Results

All data and analysis scripts associated with Experiment 1 are publicly available on the OSF and can be accessed at https://osf.io/m7df6/. Participants took an average of 50.07 min ($SD = 5.69$) to complete the entire experiment. The proportions of images for which participants successfully generated JOLs during the 6 s interval was 96.8% (see Table 1). Descriptive results of mean JOLs, correct response rate (i.e., the proportion of items for which response error was 0), and mean error distance (i.e., the average of response errors across trials) are summarized in Table 2.

Table 1 Proportions of images for which participants made JOLs in experiments 1 and 2

	No-suppression	Suppression	Difference [95% CI]	t	*p*	Cohen's d	BF_{10}
Experiment 1	96.8% (3.9%)	--	--	--	--	--	--
Experiment 2	95.1% (5.1%)	95.7% (3.2%)	0.9% [0.1%, 1.7%]	−0.51	0.61	−0.15	0.32

Note. The second and third columns list *M* (*SD*) of the proportions of images for which participants successfully made JOLs during the limited time window

Table 2 Descriptive results in experiments 1 and 2

		JOL	no-JOL
Experiment 1			
	Mean JOLs	45.75 (12.32)	--
	Correct response rate	1.6% (1.3%)	1.4% (1.2%)
	Mean error distance	70.55 (9.73)	73.86 (11.32)
Experiment 2			
Control	Mean JOLs	42.57 (10.18)	--
	Correct rate	1.0% (0.9%)	1.2% (1.0%)
	Mean error distance	71.71 (8.59)	76.61 (7.38)
Suppression	Mean JOLs	46.85 (12.62)	--
	Correct response rate	1.2% (1.1%)	0.9% (0.8%)
	Mean error distance	75.10 (6.59)	75.14 (5.59)

Aggregated modeling analyses

Figure 3a and b show the distributions of response errors in the JOL and no-JOL conditions, respectively. An aggregated mixture-modeling analysis showed that memory accessibility was greater in the JOL (P_{mem} = 0.30, 95% CrI = [0.26, 0.34]) than in the no-JOL (P_{mem} = 0.21, 95% CrI = [0.18, 0.24]) condition (see Fig. 3c). In contrast, mnemonic precision was poorer in the JOL (SD_{mem} = 30.90, 95% CrI = [26.93, 37.07]) than in the no-JOL (SD_{mem} = 18.65, 95% CrI = [15.75, 21.66]) condition (see Fig. 3d). These results imply that making JOLs reactively enhances accessibility but concurrently impairs precision of continuous color memory.

Individual-level modeling analyses

Results from individual-level mixture-modeling analyses are consistent with those from aggregated mixture-modeling analyses. Specifically, the results again showed greater memory accessibility in the JOL (*M* of P_{mem} = 0.33, *SD* = 0.15) than in the no-JOL (*M* = 0.24, *SD* = 0.13) condition, difference = 0.09, 95% CI = [0.04, 0.14], $t(24) = 3.90$, $p < .01$, Cohen's $d = 0.78$, $BF_{10} = 49.19$ (see Fig. 3e). In contrast, mnemonic precision was lower in the JOL (*M* of SD_{mem} = 36.35, *SD* = 16.32) than in the no-JOL (*M* = 21.21, *SD* = 9.63) condition, difference = 15.14, 95% CI = [8.78, 21.50], $t(24) = 4.91$, $p < .01$, Cohen's $d = 0.98$, $BF_{10} = 483.41$ (see Fig. 3f).

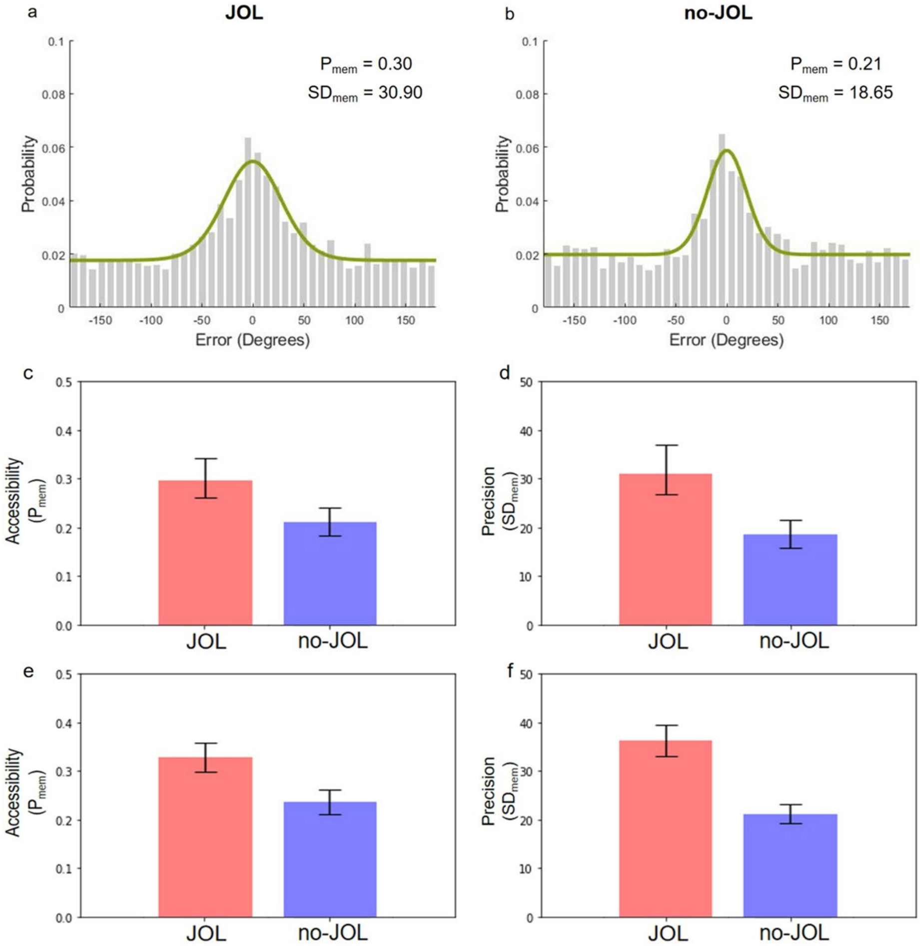

Fig. 3 Results of Experiment 1. *Note*. (**a**) Aggregate fit of response errors in the JOL condition. (**b**) Aggregate fit of response errors in the no-JOL condition. (**c**) Aggregate estimates of memory accessibility (P_{mem}); (**d**) Aggregate estimates of memory precision (SD_{mem}). In panels c and d, error bars represent 95% CrIs of the estimated parameters. (**e**) Average of individual-level parameter estimates of memory accessibility (P_{mem}); (**f**) Average of individual-level parameter estimates of memory precision (SD_{mem}). In panels e and f, error bars represent 95% within-subjects CIs

Color response distribution analyses

According to the dual-trace theory of visual memory, visual memory comprises two sources of information: visual representation and verbal label (Donkin et al., 2015). Participants' responses in the color reproduction test could stem from both memory traces (Brandimonte et al., 1992). For instance, when a given participant saw a counter shape of snake, she might recall that she had previously studied a "*red snake*", and then used such a verbal label to make a color response (i.e., selecting red color for the snake shape). Meanwhile, she might make a color response according to her visual representation of the original image viewed during the study phase (Souza & Skóra, 2017).

According to the strategy-change theory, we suspect that the requirement of making JOLs might prompt participants to use the verbal-labeling strategy to encode and retrieve color information. Specifically, during the encoding phase, when participants are asked to predict the likelihood of remembering the color of a given animal shape, they have to search for "diagnostic" cues to guide JOL formation. The task requirement of making JOLs might drive them to use the verbal-labeling strategy to encode color information (e.g., *red snake*), and then make a JOL according to the memory strength of the association between the animal shape and the categorical color label.

It has been established that precision of continuous color memory is largely dependent on the quality of visual representation, whereas retrieval accessibility heavily relies on memory strength of verbal labels (Fougnie et al., 2016; Hardman & Cowan, 2015; Palmer et al., 2015; Swan et al., 2016). For instance, in a continuous visual memory study conducted by Souza and Skóra (2017), participants were sequentially presented with colored disks. During the learning phase, participants were instructed to either verbally label the colors (referred to as the labeling condition) or repeat the syllable "bababa" aloud to prevent verbal labeling (referred to as the articulatory suppression condition). The results from a mixture-modeling analysis showed a much larger P_{mem} in the labeling than in the articulatory suppression condition, suggesting that verbal labeling does enhance verbal representation of visual information and facilitate memory accessibility.

Even though verbal labeling can enhance memory accessibility, it may concurrently overshadow precision of continuous color memory. For instance, when encountering the shape of a snake with a hue set at [200, 226, 117], utilizing the verbal-labeling strategy may make participants to verbally represent it as "*green snake*". This categorical color label could help participants access the stored information that the snake is in an approximate shade of green, but the deviation of "green" [0, 255, 0] from the actual color [200, 226, 117] could lead to impaired precision. That is, storage of the term "*green*" at the expense of the particular greenish hue should lead to a loss of memory precision, which is termed as a *verbal overshadowing effect* (Souza & Skóra, 2017).

Overall, a possible explanation for the dissociated reactivity effects of making JOLs on accessibility (positive reactivity) and precision (negative reactivity) of continuous color memory is that the requirement of making JOLs changes learners' study strategies by driving them to rely more on the verbal-labeling strategy to encode visual information, which in turn leads to enhanced accessibility and impaired precision of continuous color memory.

Differences in color response distributions were analyzed to test this possible explanation. If a given participant completely relied on continuous visual representations (rather than categorical verbal representations) to encode and recall color information, we would expect that her color responses (rather than response errors) in the color reproduction test would also roughly follow a uniform distribution on the color wheel. Otherwise, if she mainly relied on the verbal-labeling strategy to encode and recall shape colors (e.g., *green-snake*), we expect that she would frequently select typical colors (e.g., red [255, 0, 0], green [0, 255, 0]) and her color responses in the color reproduction test would severely violate a uniform distribution. According to the strategy-change theory of JOL reactivity (i.e., the verbal-labeling explanation), we predict that participants' color responses would violate uniform distribution more severely in the JOL than in the no-JOL condition.

Figure 4a and b depict the frequency of color responses on the color wheel in the JOL and no-JOL conditions, respectively. The peaks and troughs of color response distribution

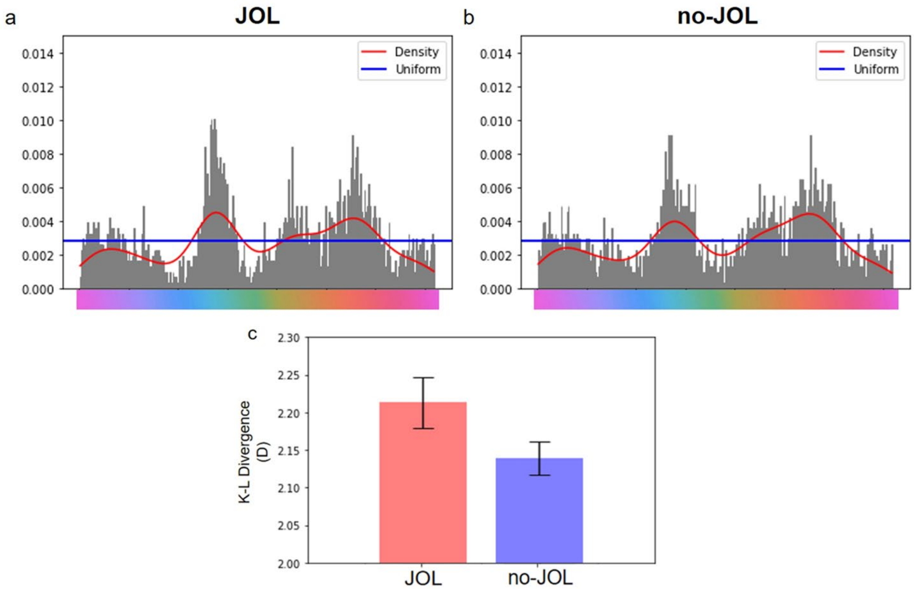

Fig. 4 Color response distributions in Experiment 1. *Note*. (**a**) Frequency distribution of color responses in the JOL condition. (**b**) Frequency distribution of color responses in the no-JOL condition. In panels a and b, the red curves represent the density curves of participants' color response distribution, and the blue straight lines represent the density line of a uniform distribution. (**c**) Average of estimates of parameter D as a function of study method. In panel c, error bars represent 95% within-subjects CIs

appear narrower in the JOL than in the no-JOL condition. To test this, we employed Kullback-Leibler divergence value (Parameter D) to quantify the degree of deviation between participants' color response distribution and a uniform distribution for each participant in each of the JOL and no-JOL conditions (Kullback & Leibler, 1951). Parameter D is a statistical measure used to assess the difference between two probability distributions. Moreno et al. (2003) used it for multimedia classification, and Dhillon et al. (2003) used it for text classification. In research of machine learning and neuroscience, parameter D is frequently employed to approximate challenging density models (Minka, 2001). Parameter D provides a way to test for strategic responding in the JOL condition.

All participant-level K-L Divergence (D) values were estimated using *NumPy toolkit* in Python 3.09. Then, a Wilcoxon rank-sum test was performed using JASP 0.19.3.0. The results showed greater parameter D in the JOL (M=2.21, SD=0.17) than in the no-JOL (M=2.14, SD=0.11) condition, Hodges-Lehmann estimate=0.07, 95% CI = [0.02, 0.13], W=262.00, z=2.68, $p<.01$, Rank-Biserial correlation=0.61, BF_{10}=6.40 (see Fig. 4c). This finding suggests that the requirement of making JOLs induces a stronger preference for using the verbal-labeling strategy to encode and recall color information (i.e., a stronger preference to use categorical color labels to represent visual information), supporting the verbal-labeling explanation.

Discussion

Experiment 1 observed dissociated reactivity effects on continuous color memory: Making JOLs enhanced accessibility but impaired precision of continuous color memory, as reflected by greater P_{mem} (i.e., greater accessibility) and greater SD_{mem} (i.e., lower precision) in the JOL than in the no-JOL condition. Furthermore, analyses on the shape of the color response distributions suggested that participants were more likely to use a verbal-labeling strategy in the JOL than in the no-JOL condition, consistent with the strategy-change hypothesis of the JOL reactivity effect.

Experiment 2

Experiment 1 provided initial evidence that soliciting JOLs reactively enhances accessibility but undermines precision of continuous color memory. Clearly, a single experiment is insufficient to make a firm conclusion. Therefore, the first aim of Experiment 2 was to replicate the reactivity findings observed in Experiment 1. To accomplish this aim, Experiment 2 included a control group, for which the stimuli and experimental procedure were the same as those in Experiment 1.

The second aim of Experiment 2 was to further test the verbal-labeling explanation. As aforementioned, Experiment 1 observed that the distribution of color responses deviated more from uniform distribution in the JOL than in the no-JOL condition, supporting the verbal-labeling explanation of the dissociated reactivity effects on accessibility and precision of continuous color memory. Going beyond Experiment 1, Experiment 2 employed an articulatory suppression procedure to further test the role of verbal labeling (i.e., strategy change) in the dissociated reactivity effects. Specifically, Experiment 2 included a suppression group, in which participants were asked to verbally repeat the syllable "bababa" aloud during the encoding phase. Numerous studies have confirmed that articulatory suppression can prevent participants from generating verbal labels (Morey & Cowan, 2004; Overkott & Souza, 2023; Sense et al., 2017; Souza & Skóra, 2017). Therefore, according to the verbal-labeling explanation, we predicted that articulatory suppression would reduce (or even eliminate) the dissociated reactivity effects on accessibility and precision of continuous color memory. Furthermore, we predicted that, in the control group, the distribution of color responses would violate the uniform distribution more severely in the JOL than in the no-JOL condition. In contrast, in the suppression group, the level of deviation from uniform distribution between the two conditions would be reduced (or even eliminated).

Methods

Participants

Akin to Experiment 1, a pilot study, with 10 participants in each of the control and suppression groups, found that the effect sizes for the interaction between group (control vs. suppression) and study method (JOL vs. no-JOL) were $\eta_p^2 = 0.147$ for P_{mem} and $\eta_p^2 = 0.227$ for SD_{mem}. Detailed results of the pilot study are reported in the Appendix. A power analysis indicated that 24 participants in each group were required to detect a significant interaction (α=0.05) for parameter P_{mem} with a power of 0.80.

Accordingly, Experiment 2 recruited 48 new participants from the participant pool at Beijing Normal University. Five extra participants were recruited as replacements for those whose data were excluded because their estimated parameters of P_{mem} or SD_{mem} deviated by more than three *SD*s from the group mean. In total, the final data came from 48 participants (M_{age} = 21.60, *SD*=2.24; 32 female), with 24 in each group. All participants provided informed consent, were tested individually in a sound-proofed cubicle, reported normal or corrected-to-normal vision, and received 50 RMB as task compensation.

Materials, design and procedure

Experiment 2 involved a 2 (group: control vs. suppression) × 2 (study method: JOL vs. no-JOL) mixed design, with group as a between-subjects factor and study method as a within-subjects factor. For both groups, the stimuli and procedure were the same as those in Experiment 1, but with only one difference in the suppression group. Specifically, before initiating the experiment, participants in the suppression group needed to practice the articulatory suppression task. The computer program repeatedly played the sound "ba" at 1 s intervals, and participants needed to try to follow along with the sound until they could keep up with the rhythm proficiently. When participants felt they could independently produce the sound in rhythm without computer's assistance, they pressed the Space bar to start the practice phase of the experiment, at which point the computer program stopped playing the sound. Throughout the entire practice and learning phases, participants in the suppression group were asked to verbally repeat the "bababa" syllable aloud to inhibit verbal labeling. An experimenter monitored from a corner of the laboratory to ensure participants in both groups followed the instructions.

Results

All data and analysis scripts associated with Experiment 2 are publicly available on the OSF and can be accessed at https://osf.io/m7df6/. The overall experiment duration for the suppression group (*M* of duration=53.49 min, *SD*=3.89) was longer than the control group (*M*=48.60 min, *SD*=5.36), difference=−4.89, 95% CI = [−7.67, −2.10], $t(46)=-3.54$, $p<.01$, Cohen' s $d=-1.02$, $BF_{10}=32.72$. This difference was primarily due to the fact that the suppression group needed to complete a practice of articulatory suppression before starting the learning task. The proportions of images for which participants successfully generated JOLs during the 6 s interval was summarized in Table 1. Importantly, there was no detectable difference in the rates of making JOLs between the control group (*M*=95.1%, *SD*=5.1%) and the suppression group (*M*=95.7%, *SD*=3.2%), difference=0.9%, 95% CI = [0.1%, 1.7%], $t(23)=-0.51$, $p=.61$, Cohen's $d=-0.15$, $BF_{10}=0.32$, suggesting that articulatory suppression produced minimal impacts on the JOL-making process. Descriptive results of mean JOLs, correct response rate, and mean error distance are shown in Table 2.

Aggregated modeling analysis

The reactivity findings observed in Experiment 1 were successfully replicated in Experiment 2's control group. Specifically, in the control group, retrieval accessibility was greater in the JOL (P_{mem} = 0.26, 95% CrI = [0.23, 0.30]) than in the no-JOL (P_{mem} = 0.18, 95% CrI = [0.16, 0.22]) condition (see Fig. 5e). In contrast, memory precision was poorer in the JOL (SD_{mem} = 28.82, 95% CrI = [24.80, 35.01]) than in the no-JOL (SD_{mem} = 19.86, 95% CrI = [17.10, 23.60]) condition (see Fig. 5f).

In contrast to the control group, the suppression group showed minimal difference in retrieval accessibility between the JOL (P_{mem} = 0.21, 95% CrI = [0.18, 0.25]) and no-JOL (P_{mem} = 0.20, 95% CrI = [0.17, 0.24]) conditions (see Fig. 5e). In the same line, there was minimal difference in mnemonic precision between the JOL (SD_{mem} = 27.05, 95% CrI = [22.75, 33.81]) and no-JOL (SD_{mem} = 23.71, 95% CrI = [20.05, 27.52]) conditions (see Fig. 5f). These findings reflect that the articulatory suppression manipulation successfully eliminated the dissociated reactivity effects on accessibility and precision of continuous color memory.

Individual-level modeling analysis

Frequentist and Bayesian mixed ANOVAs showed that, for individual-level parameter P_{mem}, there was a significant interaction between group and study method, $F(1,46)=6.09$, $p=.02$, $\eta_P^2=0.06$, BF_{incl} = 3.45 (see Fig. 5g). This interaction arose from the fact that the positive reactivity effect on memory accessibility (calculated as the difference in P_{mem} between the JOL and no-JOL conditions) was smaller in the suppression ($M<0.01$, $SE=0.02$) than in the control group ($M=0.06$, $SE=0.02$). A similar interaction was also found for individual-level parameter SD_{mem}, $F(1,46)=6.37$, $p=.01$, $\eta_P^2=0.07$, BF_{incl} = 3.91 (see Fig. 5h). This interaction arose from the fact that the negative reactivity effect on memory precision (calculated as the difference in SD_{mem}) was smaller in the suppression ($M=1.84$, $SE=2.73$) than in the control group ($M=7.51$, $SE=2.73$).

Further *t*-tests showed that, in the control group, memory accessibility was greater in the JOL (M of P_{mem} = 0.29, $SD=0.12$) than in the no-JOL ($M=0.19$, $SD=0.09$) condition, difference=0.11, 95% CI = [0.06, 0.16], $t(23)=4.51$, $p<.01$, Cohen's $d=0.92$, $BF_{10}=178.47$ (see Fig. 5g). By contrast, memory precision was poorer in the JOL (M of SD_{mem} = 34.82, $SD=14.83$) than in the no-JOL ($M=20.42$, $SD=7.64$) condition, difference=14.40, 95% CI = [7.94, 20.87], $t(23)=4.61$, $p<.01$, Cohen's $d=0.94$, $BF_{10}=224.61$ (see Fig. 5h). Overall, these results successfully replicate Experiment 1's findings and re-confirm the dissociated reactivity effects on continuous color memory.

By contrast, in the suppression group, there was little difference in memory accessibility between the JOL ($M=0.24$, $SD=0.08$) and no-JOL ($M=0.24$, $SD=0.10$) conditions, difference=0.01, 95% CI = [−0.05, 0.06], $t(23)=0.30$, $p=.77$, Cohen's $d=0.06$, $BF_{10}=0.22$ (see Fig. 5g). Similarly, there was little difference in memory precision between the JOL ($M=29.77$, $SD=13.39$) and no-JOL ($M=29.15$, $SD=16.08$) conditions, difference=0.62, 95% CI = [−8.31, 9.54], $t(23)=0.14$, $p=.89$, Cohen's $d=0.03$, $BF_{10}=0.22$ (see Fig. 5h). Overall, these results reflect that articulatory suppression successfully eliminated the dissociated reactivity effects on continuous color memory, again supporting the verbal-labeling explanation.

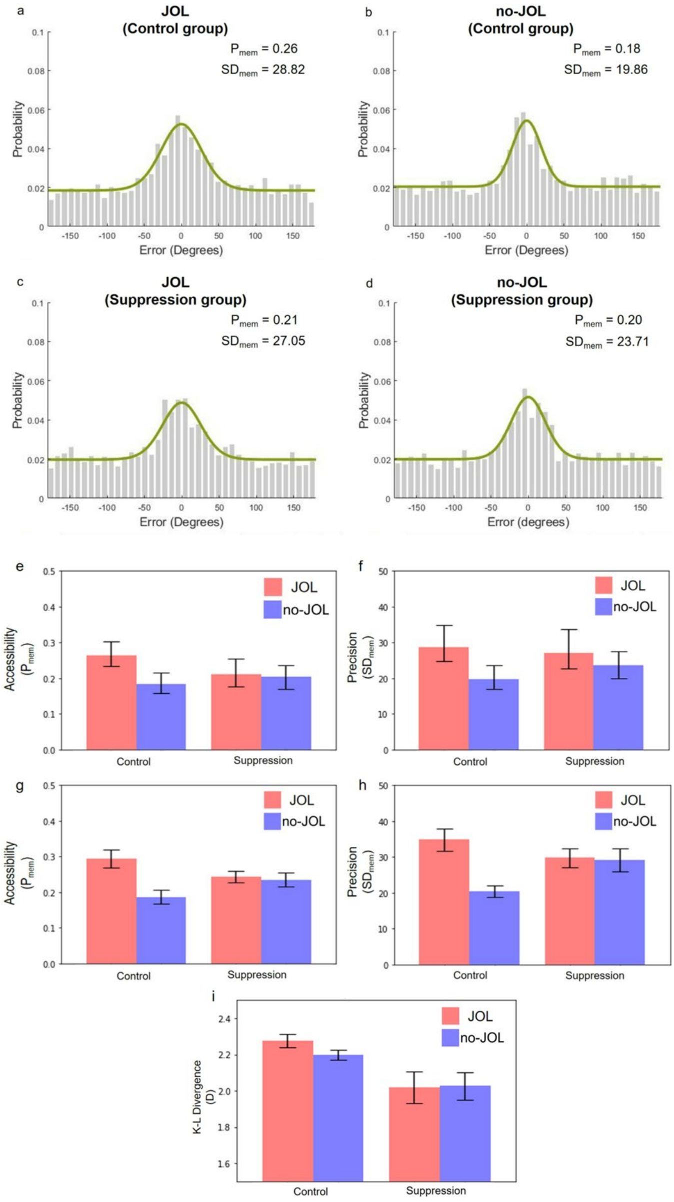

Fig. 5 Results of Experiment 2. *Note*. (**a**) Aggregate fit of response errors in the control group's JOL condition. (**b**) Aggregate fit of response errors in the control group's no-JOL condition. (**c**) Aggregate fit of response errors in the suppression group's JOL condition. (**d**) Aggregate fit of response errors in the suppression group's no-JOL condition. (**e**) Aggregate parameter estimates of P_{mem} as a function of study method and group. (**f**) Aggregate parameter estimates of SD_{mem} as a function of study method and group. In panels e and f, error bars represent 95% CrIs of the estimated parameters. (**g**) Average of individual-level parameter estimates of P_{mem} as a function of study method and group; (**h**) Average of individual-level parameter estimates of SD_{mem} as a function of study method and group. (**i**) Average of estimates of parameter D as a function of study method and group. In panels g-i, error bars represent 95% within-subjects CIs

Color response distribution analysis

Wilcoxon rank-sum tests showed that, in the control group, parameter D was greater in the JOL ($M=2.28$, $SD=0.17$) than in the no-JOL ($M=2.20$, $SD=0.13$) condition, Hodges-Lehmann estimate $=0.05$, 95% CI = [0.02, 0.11], $W=249.00$, $z=2.83$, $p<.01$, Rank-Biserial correlation $=0.66$, $BF_{10}=6.33$ (see Fig. 5i). These results replicated Experiment 1's findings and indicate that the requirement of making JOLs induces a stronger preference for using category labels to encode and retrieve color information.

By contrast, in the suppression group, there was minimal difference in D values between the JOL ($M=2.02$, $SD=0.43$) and no-JOL ($M=2.03$, $SD=0.37$) conditions, Hodges-Lehmann estimate <0.001, 95% CI = [−0.07, 0.07], $W=151.00$, $z=0.029$, $p=.99$, Rank-Biserial correlation $=0.01$, $BF_{10}=0.22$ (see Fig. 5i), indicating that the articulatory suppression manipulation almost completely eliminated the preference of using verbal labels to encode and recall color information in the JOL condition.

Discussion

The control group's results successfully replicated the main findings of Experiment 1 by showing a positive reactivity effect on accessibility and a negative reactivity effect on precision of continuous color memory. The distribution comparison results also replicated Experiment 1's findings by showing that the control group's color response distribution deviated from uniform distribution more severely in the JOL than in the no-JOL condition, supporting the verbal-labeling explanation. More importantly, the articulatory suppression manipulation, which was effective at inhibiting verbal labeling, successfully eliminated the dissociated reactivity effects on continuous color memory, again supporting the verbal-labeling explanation.

General discussion

Previous studies on JOL reactivity have generally assessed memory in a binary "all-or-none" approach. In contrast, the current research conducted the first exploration of whether making JOLs reactively alters continuous color memory measured in an "all-or-some-or-none" approach. Both Experiment 1 and Experiment 2's control group consistently demonstrated that soliciting JOLs reactively enhanced accessibility but impaired precision of continuous color memory. These findings thus extend the reactivity effect from traditional binary indicators to a continuous scale in visual memory. More importantly, the JOL reactivity effect on continuous color memory is dissociated, with a positive reactivity effect on memory accessibility and a negative reactivity effect on memory precision.

In Experiment 1, we employed a within-subjects design, which could eliminate between-subjects variability and provide greater statistical power for detecting the reactivity effect on continuous color memory (Goodhew & Edwards 2019). Notably, both within-subjects (Kubik et al., 2022; Li et al., 2024) and between-subjects (Shi et al., 2023; Zhao et al., 2023b) designs have been widely employed to explore the JOL reactivity effect. Furthermore, Rivers et al. (2021) found that the reactivity effect survived in both within- and between-subjects design experiments. More importantly, Double et al. (2018b) meta-analysis found that experimental design (within- vs. between-subjects) did not moderate the magnitude of the JOL reactivity effect. Hence, there should be little need worry about the influences of experimental design on the main findings documented here.

We propose that the strategy-change theory of JOL reactivity can reasonably explain the dissociated reactivity effects on accessibility and precision. During the encoding phase, when participants were asked to predict the likelihood of remembering the color of a given animal shape, they had to search for "diagnostic" cues to guide JOL formation. The process of cue searching might have prompted participants to use the verbal-labeling strategy to encode color information. Then, they metacognitively assessed memory strength of the association between the shape and the categorical color label, and took this as an "informative" cue to form a JOL (Castel, 2008; Dunlosky & Matvey, 2001; Koriat, 1997; Koriat & Goldsmith, 1996; Koriat & Ma'ayan, 2005; Yang et al., 2021).

The results of Experiment 1 confirmed that the color response distribution deviated from uniform distribution more severely in the JOL than in the no-JOL condition, and Experiment 2's control group successfully replicated this finding. These consistent findings indicate a stronger preference for using the verbal-labeling strategy in the JOL than in the no-JOL condition. Previous studies have suggested that verbal labeling helps learners effectively encode and recall the "approximate" color category, but that the precise information about the color is overshadowed by the verbal labeling process, leading to a verbal overshadowing effect and making it difficult for learners to recall the "precise" color (Dodson et al., 1997; Meissner et al., 2001; Souza & Skóra, 2017). Thus, the switch to a verbal labeling strategy might be responsible to the dissociative reactivity effects on accessibility and precision of continuous color memory. More importantly, Experiment 2 showed that, articulatory suppression, targeting to inhibit verbal labeling, successfully eliminated the dissociated reactivity effects on continuous color memory, further suggesting that a shift to the verbal-labeling strategy is the main cause of the dissociated reactivity effects on continuous color memory.

As aforementioned, three previous studies have tested the strategy-change theory of JOL reactivity, and provided inconsistent findings (Mitchum et al., 2016; Rivers et al., 2021; Shi et al., 2023). All of these three studies used retrospective self-reports to test the strategy-change theory. Going beyond, the current study was the first to use experimental tasks to directly and empirically test the strategy-change theory. Specifically, in Experiment 2, an articulatory suppression manipulation was introduced to inhibit participants' use of the verbal-labeling strategy in both the JOL and no-JOL conditions. Critically, articulatory suppression successfully eliminated the difference in verbal labeling between the JOL and no-JOL conditions (as reflected by no difference in parameter D between the two conditions). More importantly, after eliminating the difference in verbal labeling, both the positive reactivity effect on memory accessibility and the negative reactivity effect on memory precision disappeared. Thus, these findings directly support the strategy-change theory as an account of the JOL reactivity effect.

It should, however, be noted that the verbal labeling strategy for remembering visual colors is just one example of study strategy change induced by the requirement of making JOLs. Further research needs to test the strategy-change theory's validity in explaining the reactivity effects on memory for other types of materials (Li et al., 2021; Shi et al., 2023; Zhao et al., 2022). As discussed above, Mitchum et al. (2016) and Rivers et al. (2021) failed to find evidence supporting the strategy-change hypothesis with a self-report method, while using a different type of question, Shi et al. (2023) found evidence consistent with this theory. It is possible that strategy change may only contribute to the reactivity effect on memory for some types of materials (e.g., visual images) but not the effect on memory for other types of materials (e.g., word pairs). Future research is encouraged to further test this possibility.

Putting theoretical implications aside, the documented findings also bear implications for guiding the design of future metacognition research. Previous studies frequently used JOLs as a measurement tool for assessing people's metamemory monitoring ability (Besken, 2016; Rhodes, 2016; Yang et al., 2021; Zhao et al., 2023b). However, the reactivity effect suggests that JOL accuracy may be a biased measure of people's metamemory monitoring ability (Double & Birney, 2018a; Li et al., 2021; Rivers et al., 2021; Senkova & Otani, 2021). Experiments 1 and 2 consistently confirmed the existence of the dissociated JOL reactivity effects on continuous color memory. Hence, the memory-metamemory relationship observed in the JOL condition is subject to interference from the reactivity effect, rendering the conclusions of the memory-metamemory relationship observed in the JOL condition not necessarily applicable to the relationship in the standard no-JOL condition.

Notably, some recent studies on the JOLs reactivity effect have explored memory by categorizing it into recollection and familiarity processes (Maxwell & Huff, 2024; Zheng et al., 2024). Memory recollection and memory familiarity are commonly assessed using the remember/know procedure (Cohen et al., 2017), which measures the proportion of recollection-based and familiarity-based components in recognition memory. In the remember/know procedure, participants first judged whether each item was "old" or "new." For items judged as "old," they then indicated whether their recognition was based on "remembering" (reflecting recollection) or "knowing" (reflecting familiarity). Maxwell and Huff (2024) found that making JOLs enhanced recollection-based and facilitated familiarity-based recognition of related word pairs. Furthermore, Zheng et al. (2024) found that making JOLs increased both recollection and familiarity of studied words in a recognition test. However, the dissociable reactivity effects of making JOLs on recollection and familiarity were not observed by either Maxwell and Huff (2024) and Zheng et al. (2024), because the memory classification approach in the current research fundamentally differs from the recollection/familiarity perspective. That is, memory accessibility refers to the probability of successful retrieval, while memory precision refers to the fidelity of stored information. In contrast, recollection involves retrieval of contextual details about a past event, whereas familiarity pertains to recognition of an item without specific details (Eichenbaum et al., 2007; Yonelinas, 2002). In terms of measurement, accessibility and precision are typically assessed through response error in continuous color memory tasks, such as color reproduction, estimated using probabilistic mixture models. Recollection and familiarity are often measured via the remember/know procedure, where participants classify their recognition as either based on detailed recollection or feelings of familiarity (Cohen et al., 2017; Tulving, 1985). This distinction highlights the unique nature of both memory classification frameworks, underscoring different aspects of memory performance and retrieval processes.

Moreover, previous studies generally determined the nature of the reactivity effect about whether it is positive (Li et al., 2021; Witherby & Tauber, 2017; Zhao et al., 2022) or negative (Mitchum et al., 2016) based on a single measure of final test performance. The current study expands this by dividing memory performance into two distinct components: accessibility and precision. It was found that making JOLs enhances accessibility but comes at the cost of losing precision of continuous color memory. The corresponding implication is that future reactivity research should bear in mind that "memory performance" may encompass different components, and the JOL reactivity effects on different components may be dissociated (Zhao et al., 2023b). Such dissociated reactivity effects, varying in direction and strength, ultimately combine to produce the overall reactivity impact of JOLs on memory performance (such as recall or recognition). It is even possible that opposing JOL reactivity effects might result in a deceptive appearance of no reactivity effect (Schäfer & Undorf, 2023). Therefore, it is imperative for future research to not only examine the reactive impact of making JOLs on overall memory performance, but also to rigorously scrutinize the reactivity effects on specific components of memory.

Concluding remarks

Soliciting JOLs reactively enhances accessibility but concurrently impairs precision of continuous color memory. The strategy-change theory (more specifically, the verbal-labeling explanation) is a viable explanation of these dissociated effects.

Appendix

Details of the pilot results in Experiments 1 and 2 are shown in Table A1, which lists *M* and *SD* of P_{mem}, SD_{mem}, parameter D in the JOL and no-JOL conditions, respectively.

Table A1 Descriptive results of the pilot data in experiments 1 and 2

		JOL	no-JOL
Experiment 1	P_{mem}	0.25 (0.07)	0.19 (0.15)
	SD_{mem}	32.66 (14.26)	18.77 (7.34)
	Parameter D	2.27 (0.19)	2.20 (0.13)
Experiment 2			
Control	P_{mem}	0.29 (0.12)	0.19 (0.09)
	SD_{mem}	34.82 (14.83)	20.42 (7.64)
	Parameter D	2.28 (0.17)	2.20 (0.13)
Suppression	P_{mem}	0.24 (0.08)	0.24 (0.10)
	SD_{mem}	29.77 (13.39)	29.15 (16.08)
	Parameter D	2.02 (0.43)	2.03 (0.37)

Acknowledgements This research was supported by Beijing Philosophy and Social Science Foundation (24DTR067).

Declarations

Competing interests The authors declare that they have no conflict of interest.

References

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*, 693–712. https://doi.org/10.1007/s10648-020-09556-8

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 7.1–711. https://doi.org/10.1167/9.10.7

Berens, S. C., Richards, B. A., & Horner, A. J. (2020). Dissociating memory accessibility and precision in forgetting. *Nature Human Behaviour*, *4*, 866–877. https://doi.org/10.1038/s41562-020-0888-8

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. https://doi.org/10.1037/xlm0000246

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. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, *24*, 981–990. https://doi.org/10.1177/0956797612465439

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Verbal recoding of visual stimuli impairs mentalimagetransformations. *Memory and Cognition*, *20*, 449–455. https://doi.org/10.3758/BF03210929

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory and Cognition*, *36*, 429–437. https://doi.org/10.3758/MC.36.2.429

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*, 1581–1601. https://doi.org/10.1037/xlm0000395

Cooper, R. A., Richter, F. R., Bays, P. M., Plaisted-Grant, K. C., Baron-Cohen, S., & Simons, J. S. (2017). Reduced hippocampal functional connectivity during episodic memory retrieval in autism. *Cerebral Cortex (New York N Y 1991)*, *27*, 888–902. https://doi.org/10.1093/cercor/bhw417

Dhillon, I. S., Mallela, S., & Kumar, R. (2003). A divisive information theoretic feature clustering algorithm for text classification. *The Journal of Machine Learning Research*, *3*, 1265–1287. https://doi.org/10.5555/944919.944973

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition*, *25*, 129–139. https://doi.org/10.3758/BF03201107

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin and Review*, *22*, 170–178. https://doi.org/10.3758/s13423-014-0675-5

Double, K. S., & Birney, D. P. (2018a). Reactivity to confidence ratings in older individuals performing the Latin square task. *Metacognition and Learning*, *13*, 309–326. https://doi.org/10.1007/s11409-018-9186-5

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

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*, 1180. https://doi.org/10.1037/0278-7393.27.5.1180

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

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. https://doi.org/10.3758/BF03193146

Fougnie, D., Cormiea, S. M., Kanabar, A., & Alvarez, G. A. (2016). Strategic trade-offs between quantity and quality in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1231–1240. https://doi.org/10.1037/xhp0000211

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008

Halamish, V., & Undorf, M. (2023). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *49*, 547–556. https://doi.org/10.1037/xlm0001174

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*, 325–347. https://doi.org/10.1037/xlm0000031

Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin and Review*, *20*, 318–325. https://doi.org/10.3758/s13423-012-0340-9

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short-and long-term visual memory. *Journal of Experimental Psychology: Human Perception Performance*, *30*, 519–537. https://doi.org/10.1037/0096-1523.30.3.519

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 and Review*, *25*, 2356–2364. https://doi.org/10.3758/s13423-018-1463-4

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. https://doi.org/10.1177/0956797610385359

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. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*, 478–492. https://doi.org/10.1016/j.jml.2005.01.001

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

Korkki, S. M., Richter, F. R., Jeyarathnarajah, P., & Simons, J. S. (2020). Healthy ageing reduces the precision of episodic memory retrieval. *Psychology and Aging*, *35*, 124–142. https://doi.org/10.1037/pag0000432

Kubik, V., Koslowski, K., Schubert, T., & Aslan, A. (2022). Metacognitive judgments can potentiate new learning: The role of covert retrieval. *Metacognition and Learning*, *17*, 1057–1077. https://doi.org/10.1007/s11409-022-09307-w

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*, 79–86. https://www.jstor.org/stable/2236703

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 and Cognition*, *50*, 1061–1077. https://doi.org/10.3758/s13421-021-01258-y

Li, B., Shanks, D. R., Zhao, W., Hu, X., Luo, L., & Yang, C. (2024). Do changed learning goals explain why metamemory judgments reactively affect memory? *Journal of Memory and Language*, *136*, 104506. https://doi.org/10.1016/j.jml.2024.104506

Liu, S., Briscoe, J., & Kent, C. (2023). Retrieval-induced forgetting of spatial position depends on access to multiple shared features within categories. *Memory and Cognition*, *51*, 1090–1102. https://doi.org/10.3758/s13421-022-01384-1

Maxwell, N. P., & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning*, *17*, 589–625. https://doi.org/10.1007/s11409-022-09301-2

Maxwell, N. P., & Huff, M. J. (2024). Judgment of learning reactivity reflects enhanced relational encoding on cued-recall but not recognition tests. *Metacognition and Learning*, *19*, 189–213. https://doi.org/10.1007/s11409-023-09369-4

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory and Cognition*, *29*, 176–186. https://doi.org/10.3758/BF03195751

Minka, T. P. (2001). *A family of algorithms for approximate Bayesian inference* (Doctoral dissertation, Massachusetts Institute of Technology). https://dspace.mit.edu/bitstream/handle/1721.1/86583/48118181-MIT

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. https://doi.org/10.1037/a0039923

Moreno, P., Ho, P., & Vasconcelos, N. (2003). A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In *Proceedings of the 16th International Conference on Neural Information Processing Systems* (pp. 1385–1392). https://doi.org/10.5555/2981345.2981516

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin and Review*, *11*, 296–301. https://doi.org/10.3758/BF03196573

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

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

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation*. *Elsevier*, *26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5

Nilakantan, A. S., Bridge, D. J., Gagnon, E. P., VanHaerents, S. A., & Voss, J. L. (2017). Stimulation of the posterior Cortical-Hippocampal network enhances precision of memory recollection. *Current Biology*, *27*, 465–470. https://doi.org/10.1016/j.cub.2016.12.042

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*, 341–364. https://doi.org/10.1037/a0018926

Overkott, C., & Souza, A. S. (2023). The fate of labeled and nonlabeled visual features in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *49*, 384–407. https://doi.org/10.1037/xhp0001089

Palmer, J., Boston, B., & Moore, C. M. (2015). Limited capacity for memory tasks with multiple features within a single object. *Attention Perception and Psychophysics*, *77*, 1488–1499. https://doi.org/10.3758/s13414-015-0909-2

Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. *The Oxford Handbook of Metamemory*, *1*, 65–80. https://psycnet.apa.org/record/2016-05591-004

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

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*, 1342–1353. https://doi.org/10.1080/09658211.2021.1985143

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language*, *48*(1), 195–206. https://doi.org/10.1016/S0749-596X(02)00524-7

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin and Review*, *11*, 131–136. https://doi.org/10.3758/BF03206472

Schäfer, F., & Undorf, M. (2023). On the educational relevance of immediate judgment of learning reactivity: No effects of predicting one's memory for general knowledge facts. *Journal of Applied Research in Memory and Cognition Advance Online Publication*. https://doi.org/10.1037/mac0000113

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory and Cognition*, *49*, 955–967. https://doi.org/10.3758/s13421-020-01133-2

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, *49*, 853–862. https://doi.org/10.3758/s13428-016-0741-1

Shi, A., Xu, C., Zhao, W., Shanks, D. R., Hu, X., Luo, L., & Yang, C. (2023). Judgments of learning reactively facilitate visual memory by enhancing learning engagement. *Psychonomic Bulletin and Review*, *30*, 676–687. https://doi.org/10.3758/s13423-022-02174-1

Smith, A. R. (1978). Color gamut transform pairs. *ACM Siggraph Computer Graphics*, *12*, 12–19. https://doi.org/10.1145/965139.807361

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. https://doi.org/10.1037/a0038388

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

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

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*, 91–98. https://doi.org/10.1167/13.10.9

Sutterer, D. W., & Awh, E. (2016). Retrieval practice enhances the accessibility but not the quality of memory. *Psychonomic Bulletin and Review*, *23*, 831–841. https://doi.org/10.3758/s13423-015-0937-x
Swan, G., Collins, J., & Wyble, B. (2016). Memory for a single object has differently variable precisions for relevant and irrelevant features. *Journal of Vision*, *16*(1). -32.12
Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *Quarterly Journal of Experimental Psychology*, *65*, 1376–1396. https://doi.org/10.1080/17470218.2012.656665
Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, *34*, 836–847. https://doi.org/10.1037/pag0000376
Tulving, E. (1985). Memory and consciousness. *Canadian Psychology-Psychologie Canadienne*, *26*, 1–12. https://doi.org/10.1037/h0080017
Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*, 8780–8785. https://doi.org/10.1073/pnas.1117465109
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. https://doi.org/10.1016/j.jarmac.2017.08.004
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
Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S. 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. https://doi.org/10.1007/s11409-020-09254-4
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517. https://doi.org/10.1006/jmla.2002.2864
Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*, 800–832. https://doi.org/10.1037/0033-2909.133.5.800
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. https://doi.org/10.1038/nature06860
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*, 405–417. https://doi.org/10.1111/cdev.13689
Zhao, W., Xu, M., Xu, C., Li, B., Hu, X., Yang, C., & Luo, L. (2023a). Judgments of learning following retrieval practice produce minimal reactivity effect on learning of education-related materials. *Journal of Intelligence*, *11*(190.1). -190.15
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*, 549–566. https://doi.org/10.1007/s11409-023-09340-3
Zheng, J., Li, B., Zhao, W., Su, N., Fan, T., Yin, Y., Hu, Y., Hu, X., Yang, C., & Luo, L. (2024). Soliciting judgments of learning reactively facilitates both recollection-and familiarity-based recognition memory. *Metacognition and Learning*, *19*, 609–633. https://doi.org/10.1007/s11409-024-09382-1