The Forward Effect of Judgements of Learning on Memory and Transfer in Inductive Learning

Draft version (May 02, 2025). This paper has not been peer reviewed.

Christian Ritter

University Würzburg, Germany

Hannah Hausman

University of California Santa Cruz, USA

Robert Gaschler

FernUniversität Hagen, Germany

Simon P. Tiffin-Richards

IPN - Leibniz Institute for Science and Mathematics Education, Kiel, Germany

Veit Kubik

University Würzburg, Germany

Author Note

Correspondence to be sent to Christian Ritter, Department of Psychology IV, Röntgenring 10, University of Würzburg, 97070 Würzburg, Germany. Phone: +49 (0)152 341 92218; Email: christian.ritter@uni-wuerzburg.de. The work on this article was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for the Research Unit “Lasting Learning: Cognitive mechanisms and effective instructional implementation” (Grant FOR 5254/1, project number 450142163).

Abstract

Making judgments of learning (JOLs) can directly affect learning outcomes. The present study investigated whether providing JOLs during inductive learning tasks improves learning of new material (forward effect) and whether feedback moderates these effects. Participants learned the painting styles of different artists (Experiment 1) or different rock types (Experiment 2) across two study phases separated by an interim learning task, and then completed a transfer posttest (classifying new exemplars) and memory posttest (classifying previously-studied exemplars). In Experiment 1, the interim learning tasks of overt retrieval and cue-only JOLs (based on the painting without the artist name) improved future inductive learning compared to restudy, whereas cue–target JOLs did not (painting and artist name shown). Cue-only JOLs also produced response-time patterns consistent with retrieval-based processing, and self-reported retrieval use predicted their forward benefit. Experiment 2 replicated the beneficial forward effect of cue-only JOLs over restudy with different materials and found that providing item-by-item feedback did not change the effect. Our results suggest that cue-only JOLs, but not cue–target JOLs, enhancing future inductive learning of natural visual categories, likely through metacognitively controlled, covert retrieval processes.

Keywords: Judgments of learning, Retrieval, Inductive learning, Feedback

Inductive learning is a ubiquitous form of learning to acquire knowledge without the need of formal instruction. It plays an important role in education, enabling students to derive general principles from specific examples (Holland et al., 1989). For example, in language learning, students encounter sentences in context and infer grammatical rules (Ellis, 2008). In computer science, programmers often work through examples of code to understand algorithms and generalize principles of coding logic (Guzdial & Ericson, 2016). The present study focuses on inductive learning of visual natural categories, which are common across levels of schooling. Examples range from primary school children learning types of clouds, to art history majors learning artists' styles, and medical students learning diagnoses from X-rays. Inductive learning is more than just memorizing examples; it involves synthesizing observations to extract patterns, making it a critical skill for effective problem solving and knowledge transfer across domains.

Researchers have devoted considerable effort to investigating the mechanisms of inductive learning (Holland et al., 1989; Murphy, 2004). Our goal was not to test or differentiate among theories of inductive category learning, but rather, to test promising learning tasks and gain insight into the mechanisms through which they might improve inductive learning. We investigated the potential beneficial down-stream effects of *retrieval practice* (searching memory for previously learned information) and providing *metacognitive judgments* (self-assessing understanding) on inductive learning of natural categories.

Effects of Retrieval Practice on Inductive Learning

Research has consistently shown that actively retrieving information from memory improves retention of the retrieved material, a phenomenon known as the (backward) *testing effect*. The testing effect has been demonstrated with a range of materials, from foreign-language translations to maps, in both laboratory and authentic classroom settings (Adesope et al., 2017; Carpenter et al., 2022; Rowland, 2014). The backward testing effect extends to inductive learning of natural categories. For instance, Jacoby et al. (2010) asked participants to study pictures and names of bird families and then either restudy the pairs or attempt to retrieve the family name from the picture alone. On a later test, those who practiced retrieval correctly identified more previously studied and novel exemplars. Thus, retrieval practice enhances both memory and transfer (see also Cho & Powers, 2019; Kang et al., 2023).

Retrieval practice of previously presented information has also been shown to facilitate learning of new material studied afterward. This *forward testing effect* (FTE) has been observed with single words, key term definitions, texts, videos (Chan, Meissner, & Davis, 2018; Yang et al., 2018, 2021), and inductive learning of natural categories (Ha & Lee, 2024; Kang et al., 2023; Lee & Ahn, 2018; Lee & Ha, 2019; Yang & Shanks, 2018). For example, Lee and Ahn (2018) asked participants to study paintings by several artists and then either restudy them or take an interim test. After studying new artists in a second block, participants took a transfer test, identifying new paintings by the second-block artists; participants who completed the interim test outperformed those who had restudied. Thus, explicitly testing memory, which we refer to as *overt retrieval*, is a promising strategy for improving inductive learning of both previously and subsequently studied material.

Retrieval-Based Effects of Metacognitive Judgments

The present research examined when and how providing Judgments of Learning (JOLs) may produce forward effects similar to overt retrieval. JOLs are often given on a 0–100% percent scale and involve participants estimating their likelihood of recalling a specific item on a future test (*item-specific* JOLs), correctly answering a particular type of question (*category-level* JOLs), or performing well overall (*aggregate* JOL; for a review, see Rhodes, 2016). Like overt retrieval, JOLs can have similar beneficial forward and backward effects on learning outcomes, which is sometimes referred to as *positive JOL reactivity,* although some studies report *null* or even negative effects (Double & Birney, 2019; Double et al., 2018;

Ingendahl et al., 2025; Mitchum et al., 2016; Zhao et al., 2023). A notable example of the backward benefit of JOLs on memory comes from a study by Soderstrom et al. (2015). Participants studied word pairs (e.g., *loaf–bread*) and half of participants also provided a JOL for each pair, rating the likelihood they would remember the target on a cued-recall test. Providing JOLs improved memory for the target words (for visual materials, see Shi et al., 2023; for a meta-analysis, see Ingendahl et al., 2025). Kubik and colleagues (2022) found that providing JOLs can improve memory not just for the judged material, but also for new material presented afterwards. The present research focused on this forward effect of JOLs on learning.

A leading hypothesis is that JOLs produce forward and backward effects on learning, in part, by inducing *covert retrieval*–that is, causing learners to search their memory for the target information to judge the likelihood they will be able to recall on a later test. JOLs can trigger covert retrieval attempts (e.g., Dunlosky & Nelson, 1992), and covert retrieval has been shown to produce a backward testing effect compared to restudy, although slightly smaller in magnitude than overt retrieval (for a meta-analysis, see Yu et al., 2025). Covert retrieval has also been shown to enhance subsequent new learning, with reports of a similar-sized FTE from overt and covert retrieval relative to restudy (Carvalho et al., 2023; Kubik et al., 2022).

Not all types of metacognitive judgments are presumed to induce covert retrieval, however. JOLs induce covert retrieval attempts especially when only partial information about the judged material is present (e.g., *loaf*–___; Carvalho et al., 2023; Dunlosky & Nelson, 1992; Kubik et al., 2022; Lee & Ha, 2019; Soderstrom et al., 2015; but see Zhang et al., 2024). We refer to these as *cue-only JOLs*. In contrast, when complete information is present (e.g., *loaf–bread*), these *cue–target JOLs* can be based on item familiarity without engaging in retrieval (e.g., Dunlosky & Nelson, 1992). Thus, cue-only JOLs should produce larger forward effects on new learning compared to cue–target JOLs and restudy due to a greater amount of covert retrieval. In particular, the present study tested this hypothesis using item-level JOLs, which are more likely to engage retrieval because they require monitoring of features of individual exemplars. In contrast, aggregate JOLs tend to rely on broader heuristics such as overall familiarity or fluency (Dunlosky & Nelson, 1992; Koriat, 1997) and may therefore be less sensitive to the retrieval-based mechanisms that differentiate cue-only from cue–target judgments and their forward effects on new learning. Indeed, prior research suggests that item-level and aggregate JOLs can produce different reactivity effects (Chang & Brainerd, 2024; Maxwell, 2025).

The hypothesis that cue-only JOLs produce a larger forward effect on new learning than cue–target JOLs has been supported, at least with research using relatively simple verbal learning materials. Kubik et al. (2022) had participants study five lists of words. After studying each of the first four lists, participants restudied the words (e.g., *elephant*), overtly retrieved the words (e.g., *ele*_____), or made JOLs (*How likely are you to remember this word on the test in a few minutes?*). Either the complete word (e.g., *elephant*), akin to a cue–target JOL, or only the word stem (e.g., *ele*_____), akin to a cue-only JOL, was presented with the JOL prompt. After studying List 5, all participants took a free recall test on this final list. Unlike complete-word JOLs, making word-stem JOLs after Lists 1–4 improved free recall for the final list of words compared to restudy, and the benefit was similar in magnitude to overt retrieval. The results suggest that the act of assessing one's prior learning in the absence of complete information engages covert retrieval mechanisms that facilitate new learning.

Based on the hypothesis that cue-only JOLs encourage covert retrieval, there are multiple theoretical accounts for the FTE that could be extended to explain why cue-only JOLs but not cue–target JOLs enhance new learning (for reviews, see Chan, Meissner &

Davis, 2018; Yang et al., 2018). Context accounts suggest that retrieval (covert or overt) produces a greater shift in mental context than nonretrieval-based learning tasks, thereby reducing the buildup of proactive interference and improving recall for subsequently presented material. Resource-related theories posit that encoding processes become less effective as more material is studied and interim retrieval opportunities help restore resources for subsequent encoding. Other explanations invoke more volitional processes involving metacognition (Cho et al., 2017). Attempting to retrieve previously studied material involves effort or retrieval failures that one would not experience if the complete cue–target information were present, thereby motivating increased attention and effort during subsequent encoding and retrieval. Retrieval effort and failures also provide diagnostic information about the quality of one's learning, which signal a need to adapt encoding strategies going forward.

Forward Effects of Metacognitive Judgments on Inductive Learning

We hypothesized that cue-only judgments of learning (JOLs) elicit covert retrieval, thereby producing forward benefits for new inductive learning comparable to those of overt retrieval. In contrast, we expected cue–target JOLs to function more like restudy opportunities, involving little or no retrieval and therefore producing weaker forward effects. To test these predictions, we compared learning tasks that are likely to engage retrieval processes (overt retrieval and cue-only JOLs) with those less likely to do so (cue–target JOLs and restudy).

Although research on the forward testing effect (FTE) suggests that retrieval benefits subsequent learning across a range of materials, the forward effects of JOLs remain inconsistent. For example, Lee and Ha (2019) found that interim retrieval of artist names enhanced inductive learning of new artists relative to restudy, whereas item-by-item cue–target JOLs did not. Yet in a second experiment, they reported forward benefits of cue-only JOLs comparable to retrieval practice, but with a notably different design. Participants provided just one cue-only JOL per artist based solely on the artist's name (e.g., "How likely are you to recognize new paintings by Bruno Pessani?"), without viewing any paintings. In contrast, participants in the retrieval practice condition attempted to identify one painting by each of the previously studied artists. The difference in the stimulus used to prompt cue-only JOLs versus overt retrieval leaves open questions about the extent to which cue-only JOLs function like retrieval practice. In contrast, Ha and Lee (2024) observed no forward benefit of cue-only JOLs compared to cue–target JOLs or restudy, which is somewhat difficult to interpret because no overt retrieval condition was included for comparison. These discrepancies underscore the need for a more systematic investigation of how different types of JOLs affect new learning and the role of covert retrieval in any such forward effects.

The Present Research

In two preregistered experiments, we addressed these gaps in research on the forward effects of JOLs in three key ways. First, we systematically compared the full set of theoretically-relevant learning tasks that differ in their likelihood of engaging retrieval processes: overt retrieval, cue-only JOLs, cue–target JOLs, and restudy. Including this complete set of conditions is essential for isolating retrieval's role in the forward effects of different types of JOLs.

Second, we complemented these experimental manipulations with direct measures of retrieval engagement, asking participants to self-report the extent to which they attempted retrieval during the learning task and measuring their time on task. Prior research on JOL reactivity has inferred retrieval indirectly from task format, but none has measured it directly, to our knowledge. This omission is critical because JOLs unfold in two stages: An initial familiarity assessment followed by an attempt to retrieve the target information from memory, but only when familiarity is moderate (Metcalfe & Finn, 2008; Son & Metcalfe, 2005). With perceptually rich stimuli like category paintings or pictures of butterflies, JOLs may be driven largely by familiarity (Fazendeiro et al., 2005), meaning cue-only JOLs do not always elicit as thorough of retrieval as assumed. As a result, retrieval efforts that stop prematurely will fail to support future learning—a prediction known as the *truncated search hypothesis* (Yu et al., 2025). Measuring participants' reported use of retrieval therefore provides crucial evidence about the extent to which retrieval processes underlie forward effects of JOLs.

Finally, we assessed new learning with both a memory test (targeting retention of studied exemplars) and a transfer test (assessing generalization to new exemplars). Although the forward benefits of overt retrieval are well established, we know little about whether covert retrieval produces forward effects on transfer. In fact, only one prior study has tested the forward effects of covert retrieval prompts, and it focused on simple word pairs and verbatim recall (Carvalho et al., 2023). Most research on JOLs and inductive learning, by contrast, has measured new learning solely through transfer performance (Ha & Lee, 2024; Lee & Ha, 2019; but see Wang et al., 2023, who only assessed memory), which makes it unclear whether any null effects of JOLs reflect a failure to enhance new learning altogether or a more specific failure to support transfer. If cue-only JOLs encourage retrieval of previously learned items, they may bias participants toward item-specific memorization strategies that benefit later memory tests without necessarily supporting the abstraction needed for transfer (McDaniel et al., 2018).

Participants were university students who studied paintings by artists (Experiment 1) or pictures of rock types (Experiment 2). They studied multiple exemplars from half the artists or rock categories in Phase A, completed an interim learning task related to Phase A exemplars, then studied exemplars from the other half of artists or rock categories in Phase B. Finally, participants took tests on all the categories. The learning tasks included restudy (Experiments 1 and 2), cue–target JOLs (Experiment 1), cue-only JOLs (Experiments 1 and 2), and overt retrieval (Experiment 1). Experiment 2 also manipulated whether feedback was provided after cue-only JOLs and collected different self-regulated learning measures (effort, motivation, Phase B study times) to test different theoretical explanations of how retrieval processes might promote new learning from JOLs.

Together, these design features allow us to test our central theoretical claim: that cue-only JOLs enhance new inductive learning because they engage retrieval processes. By manipulating the degree of retrieval afforded by the learning task, asking participants to report the extent of their retrieval, and assessing forward effects on both memory and transfer in natural category inductive learning, the present, preregistered experiments provide the strongest test to date of the hypothesis that covert retrieval underlies forward effects of JOLs and help clarify why previous studies have reached mixed conclusions.

Experiment 1

Experiment 1 tested the hypothesis that cue-only JOLs evoke covert retrieval attempts and therefore overt retrieval and cue-only JOLs enhance new inductive learning compared to cue–target JOLs and restudy. If covert retrieval during cue-only JOLs is sufficient to drive forward effects, then cue-only JOLs should be as effective as overt retrieval, and if metacognitive judgments do not enhance learning without retrieval, cue–target JOLs should perform no better than restudy. To test these preregistered predictions involving the four types of learning tasks, we specified three *a priori* orthogonal contrasts, which therefore fully capture all possible mean differences in learning between conditions. We predicted that learning tasks with a high probability of engaging retrieval (overt retrieval and cue-only JOLs) would enhance new inductive learning compared to low-retrieval-probability tasks (cue–target JOLs and restudy), whereas overt retrieval and cue-only JOLs would not differ significantly, nor would cue–target JOLs and restudy. Finally, exploratory analyses explored whether self-reported use of the retrieval during the learning task moderated the forward effects of cue-only JOLs. We made these predictions for both memory and transfer posttests.

The preregistration is available at the Open Science Framework: https://osf.io/ckhg2/?view_only=68f426dd0d8c43cb93b595b7b37a5bed, as are the materials, data, and analysis scripts [anonymized for peer review, requiring separate links]: https://osf.io/3qp5s/?view_only=1a55a443b3eb44a396fa01edd4168ed4.

Method

Participants

G*Power (version 3.1.9.2; Faul et al., 2007, 2013) was used to calculate the required sample size for an a priori reliable FTE or forward effect of JOLs. Using $\alpha = 0.05$, 90% power, and an effect size between $d = 0.93$ and 1.33 (cf. Lee & Ahn, 2018; Yang et al., 2018), $n = 15$ to 32 per group would be sufficient to detect group differences. We aimed for $n = 40$ participants per group ($N = 160$, total) to account for the possibility that the forward effect of cue-only JOLs may be somewhat smaller than the effect of overt retrieval (Lee & Ha, 2019).

Participants were recruited through Moodle, Unipark, and SONA from FernUniversität in Hagen and the University of Bielefeld. Approximately three-quarters of the participants were recruited from the FernUniversität in Hagen, a German distance learning university with a diverse student body in age, academic background, family status, and work history (Stürmer et al., 2018). All participants were fluent in German and received course credit. Of the 169 recruited participants, 158 were included in analyses (124 female, 33 male, 1 gender-diverse; age: $M = 29.92$ years, $SD = 11.61$, range = 18–62). Participants were excluded if they had (a) previously participated in the same study with the same learning materials ($n = 0$) or (b) used external resources during the experiment (e.g., taking notes, $n = 1$). Participants were also excluded if they showed extremely low performance on both final tests (more than 2.5 times the *MAD* below the median; Rousseeuw & Croux, 1993) coupled with reports of (c) being unfocused ($n = 4$), (d) not seriously participating ($n = 5$), or (e) speaking German less than 10 years ($n = 1$).

Design

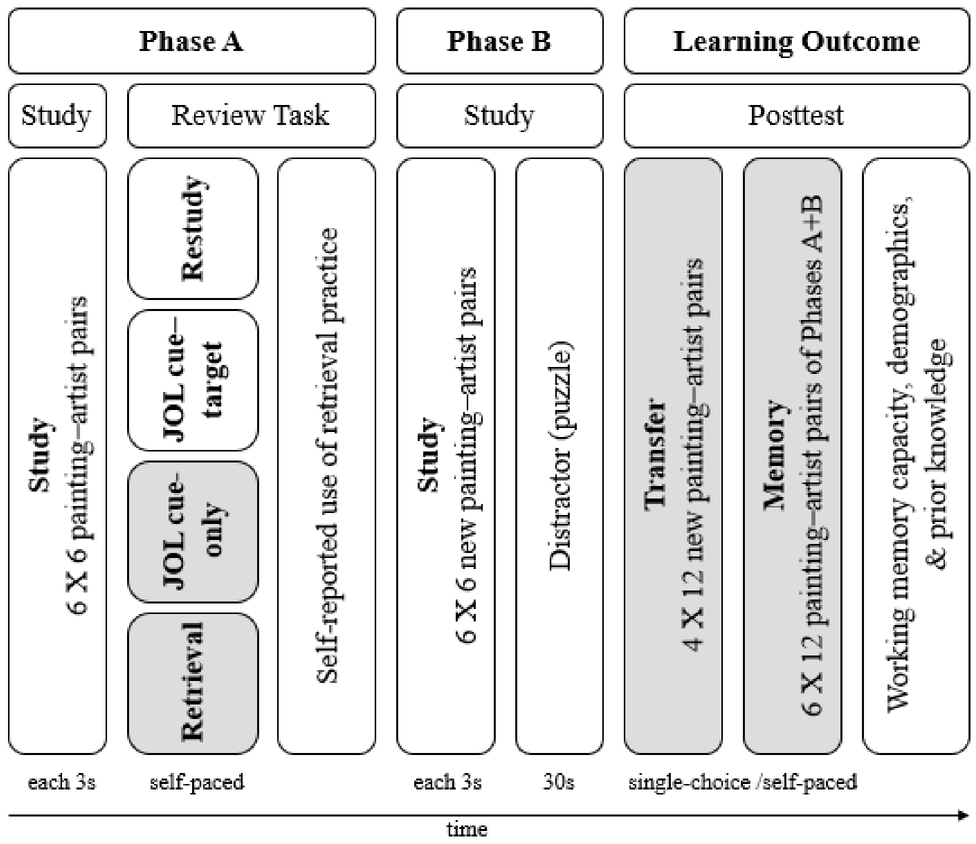
The experiment used a one-factorial, between-subjects design. Participants were randomly assigned to one of four interim learning task conditions in Study Phase A: overt retrieval, cue-only JOLs, cue–target JOLs, or restudy. The primary dependent variable was item-level classification accuracy on the transfer and memory posttests for paintings by Phase B artists. Self-reported retrieval use and learning task duration were examined as potential moderators of forward effects.

Materials and Procedure

Participants completed the study individually online via the Gorilla Experiment Builder (www.gorilla.sc, Anwyl-Irvine et al., 2020) on their personal computers at a time of their choosing. Before beginning, they received general instructions, privacy and data use information, and were told the study examined different methods of inductive learning. They were informed of the full procedure and instructed to work in a quiet, distraction-free environment. Participants were told they would study paintings by 12 artists across two phases and later identify artists' surnames for both previously studied and new paintings. The procedure is shown in Figure 1.

Figure 1

Experiment 1 Procedure



The study materials were 120 color landscape paintings (10 by each of 12 artists), adapted from Lee and Ha (2019) and Kornell and Bjork (2008).[1] Of these, 72 paintings (six per artist) served as study stimuli in Phases A (six artists) and B (six other artists); the remaining 48 paintings (four per artist) were reserved for the transfer posttest.

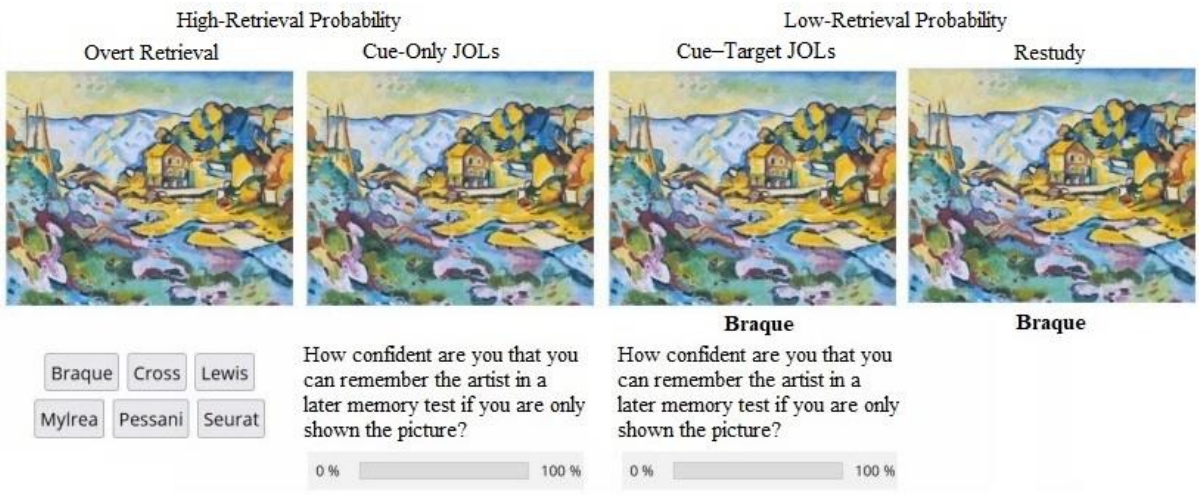
In Study Phase A, participants viewed 36 paintings (six per each of six artists), each labeled by the artist's surname. The painting–artist pairs were presented individually, in a random order, in the center of the computer screen, for 3 seconds, with a 0.5-second blank screen between pairs. Next, participants completed one of the four interim learning tasks, reviewing the same 36 paintings from Study Phase A in a new random order (Figure 2). In *overt retrieval*, participants identify the artist's surname from an alphabetized list of six names. No feedback was provided. For the *cue-only JOL* task, participants were also only

[1] The paintings were made available by Kornell and Bjork (2008) via their website (https://sites.williams.edu/nk2/stimuli/). We would like to thank Nate Kornell for providing the digital images of the paintings, and Hee Seung Lee for providing the paintings by Emma Ciardi.

presented with the painting, but rather than selecting the artist's name, they selected their JOL response. The JOL asked participants to rate the likelihood (0–100%) that they would remember the artist's surname if presented with the same painting on a later memory posttest. *Cue–target JOLs* were almost identical, but the artist's surname was shown with the painting. Finally, *restudying* involved viewing each painting–artist pair again without responding. All learning tasks were self-paced, with a 0.5-second blank screen between trials.

Figure 2

Examples of the Four Learning Tasks in Experiment 1



Participants then studied 36 new paintings (six by each of six new artists) using the same procedure as in Phase A before a 30-second distractor task (9-piece digital puzzles). Participants first completed the transfer posttest, classifying 48 new paintings (four per artist) presented in a random order by selecting the correct surname from 12 options. The memory posttest on the 72 previously studied paintings followed using the same procedure. The transfer test was always administered first to facilitate a clean interpretation of the results for comparison with prior research on inductive category learning. Finally, participants answered demographic and data quality check questions.

Instruments and Measures

Learning outcomes were assessed on the transfer posttest of 48 unseen paintings and the memory posttest of the 72 paintings shown during Phases A and B. On each trial, participants selected the correct artist surname from a list of 12 options. Accuracy was coded at the item level (1 = correct, 0 = incorrect), which served as the dependent variable in the primary mixed-effects analyses. Accuracy was also summarized as the proportion of correct responses per participant for descriptive statistics and additional exploratory analyses.

Learning task duration was measured as the total time spent on the self-paced learning tasks, recorded in milliseconds and converted to minutes.

At the end of Study Phase A, participants rated the degree to which they used retrieval practice during the learning task on a 1–9 scale with the wording tailored slightly for each condition to reflect the task instructions. Anchors in the restudy group were phrased as 1 ("I did not try to remember the name of the artist in any of the painting–artist pairs. I just looked at the painting–artist pairs.") and 9 ("For every painting–artist pair, I tried to remember the name of the artist from the painting"). Anchors in the cue-only [and cue–target JOLs] groups were 1 ("I did not try to remember the name of the artist from the painting in any of the judgments. I just looked at the paintings [painting–artist pairs]) and 9 ("I tried to remember the name of the artist from the painting for each judgment of learning"). We did not solicit ratings in the overt retrieval condition because the task inherently required retrieval.

Participants also reported demographic information (gender, age, native language, years speaking German, and highest education level) and completed data-quality checks at the end of the session. These included ratings of distraction and focus (0 = "very little", 9 = "very much"), binary questions about serious participation, legibility of materials, technical difficulties, and use of external aids, as well as prior participation in the study.

Preregistered Analysis Plan

Our primary, preregistered analyses focused on memory and transfer posttest accuracy for Phase B artists, which we refer to as Phase B memory and Phase B transfer for brevity. We used generalized linear mixed-effects models (GLMMs) with a binomial link function because the dependent variable was accuracy on each test item (correct = 1, incorrect = 0). Separate models were fit for the memory and transfer posttests. Each model was specified as $Accuracy \sim Condition + (1|\,Participant) + (1\,|\,Stimulus)$, where *Condition* was a four-level factor. Random intercepts for participants and stimuli accounted for nonindependence of observations.

Planned contrasts were defined *a priori* to test preregistered hypotheses (Wiens & Nilsson, 2017). Our central goal was to evaluate whether cue-only JOLs function as retrieval-based learning tasks, whereas cue–target JOLs function as non-retrieval-based tasks like restudy. We therefore preregistered three orthogonal contrasts that fully partition variance in accuracy and capture all mean differences among conditions. The first contrast tested a directional hypothesis, comparing high-retrieval-probability conditions (overt retrieval and cue-only JOLs) with low-retrieval-probability conditions (cue–target JOLs and restudy), with weights of +0.5, +0.5, −0.5, and −0.5.[2] Two additional preregistered contrasts directly compared overt retrieval versus cue-only JOLs (+1, −1, 0, 0) and cue–target JOLs versus restudy (0, 0, +1, –1).

To address specific questions raised during peer review, we additionally examined overt retrieval versus restudy (+1, 0, 0, −1) and cue-only JOLs versus restudy (0, +1, 0, −1). Although these pairwise contrasts are not independent of the planned high- versus low-retrieval contrast, they are derived from our preregistered predictions. If high-retrieval conditions overall outperform low-retrieval conditions and cue-only JOLs do not differ from overt retrieval, then each high-retrieval condition should individually exceed restudy.

[2] We preregistered contrast weights of ±1 but report rescaled weights of ±0.5 for interpretability. This rescaling leaves inferential tests unchanged but makes the exponentiated estimate correspond directly to the odds ratio for the main effect of high- versus low-retrieval-probability.

Accordingly, we used one-sided tests for the high- versus low-retrieval probability, overt retrieval versus restudy, and cue-only JOLs versus restudy contrasts, reporting one-sided *p*-values and 90% lower confidence bounds. All other contrasts were two-sided and reported with 95% confidence intervals. All contrast tests were based on Wald *z*-statistics, appropriate for GLMMs in large samples (Agresti, 2013). Odds ratios and their confidence intervals are reported for interpretability.

An equivalent parameterization was $Accuracy \sim Target\ condition \times JOL\ condition\ +\ (1|\ Participant)\ +\ (1\ |\ Stimulus)$. *Target condition* was target-absent (overt retrieval, cue-only JOLs) or target-present (cue–target JOLs, restudy); *JOL condition* was JOLs (cue–target JOLs, cue-only JOLs) or no JOLs (overt retrieval and restudy). Using this parameterization, we conducted exploratory omnibus Wald χ^2 tests of the main effects and their interaction. Additional exploratory analyses are described in the Results section. Analyses were performed with R (R Core Team, 2024) using packages including *lme4* (Bates et al., 2015) and *emmeans* (Lenth, 2025).

Results

Learning Task Performance

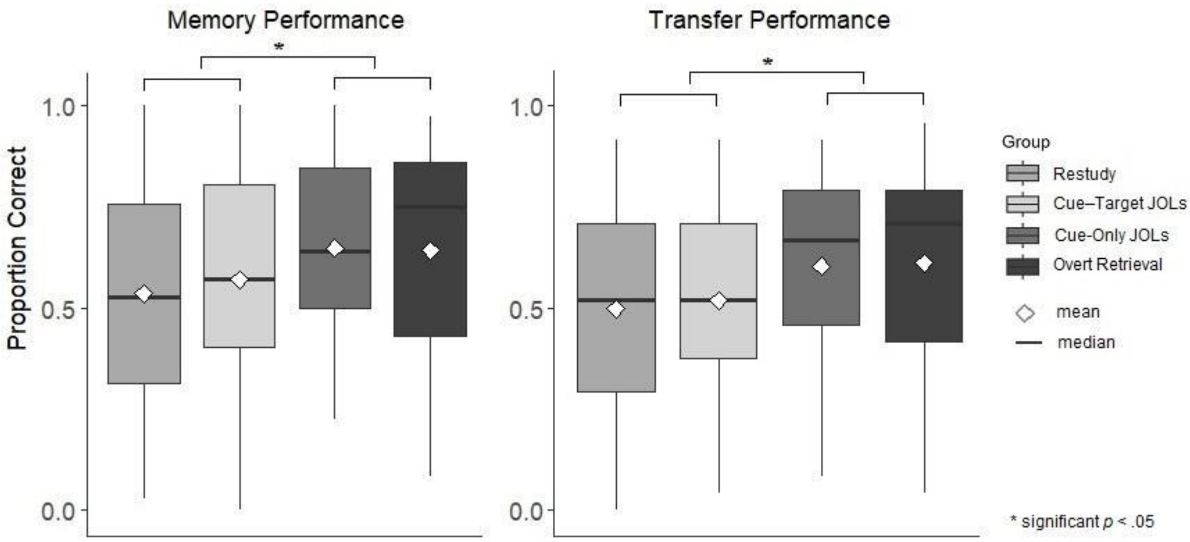
We report descriptive statistics of participants' learning task responses for completeness. Mean JOLs were similar in the cue-only (*M* = 72%, *SD* = 21%) and cue–target (*M* = 66%, *SD* = 22%) groups. The overt retrieval group was the only one tested during the learning task, and their classification accuracy (*M* = .79, *SD* = .26) was well-above chance (1/6).

Phase B Posttests

Our primary analyses examined posttest accuracy for Phase B artists (Figure 3). Omnibus Wald χ^2 tests from the GLMMs indicated a significant main effect of target condition (target-absent vs. target-present) on both memory, $\chi^2(1) = 4.86$, $p = .027$, and transfer, $\chi^2(1) = 6.64$, $p = .010$. In contrast, the main effect of JOL condition (JOL vs. no JOL) was not significant for memory, $\chi^2(1) = 0.13$, $p = .714$, or transfer, $\chi^2(1) = 0.02$, $p = .894$. The interaction between target and JOL conditions was also nonsignificant for both memory $\chi^2(1) = 0.004$, $p = .947$ or transfer accuracy, $\chi^2(1) = 0.06$, $p = .807$. Full model results and fit statistics are reported in the Supplemental Materials (Table S1). Thus, Phase B posttest accuracy was systematically related to whether the learning task afforded retrieval (target-absent) or not (target-present). Planned contrast analyses clarified the nature of these effects, as preregistered (Table 1).

Figure 3

Memory and Transfer Posttest Accuracy by Group in Experiment 1



Note. Proportion correct was aggregated by participant for the boxplots.

High- versus Low-Retrieval Probability. Consistent with our preregistered predictions, high-retrieval-probability tasks (overt retrieval and cue-only JOLs) enhanced new inductive learning more relative to low-retrieval-probability tasks (cue–target JOLs and restudy). This contrast was significant for both memory and transfer. High-retrieval groups showed higher memory accuracy ($M = .65$, $SD = .25$) than low-retrieval groups ($M = .55$, $SD = .28$), $OR = 1.76$, $p = .014$, and higher transfer accuracy ($M = .61$, SD = .24) than low-retrieval groups (*M* = .51, *SD* = .26), *OR* = 1.78, *p* = .005.

Table 1

Experiment 1 Contrasts of Posttest Accuracy for Phase B Artists

Contrast	Odds Ratio	95% CI (two-sided) / LB_{90}% (one-sided)	*z*	*p*
		Memory Performance		
High- vs. low-retrieval probability*†	1.76	> 1.27	2.21	.014
Overt retrieval vs. cue-only JOLs*	0.93	[0.45, 1.89]	–0.21	.830
Cue–target JOLs vs. restudy*	1.12	[0.55, 2.26]	0.31	.760
Overt retrieval vs. restudy	1.79	> 1.12	1.61	.054
Cue-only JOLs vs. restudy	1.93	> 1.22	1.82	.034
		Transfer Performance		
High- vs. low- retrieval probability*	1.78	> 1.33	2.58	.005
Overt retrieval vs. cue-only JOLs*	1.03	[0.55, 1.91]	0.08	.938
Cue–target JOLs vs. restudy*	1.09	[0.59, 2.01]	0.27	.788
Overt retrieval vs. restudy	1.87	> 1.25	1.99	.023
Cue-only JOLs vs. restudy	1.83	> 1.22	1.92	.027

Note. Wald *z*-tests from generalized linear mixed-effects models with random intercepts for participants and stimuli. *Indicates preregistered contrasts.

Within-Group Comparisons. To test whether cue-only JOLs were as effective as overt retrieval, we compared the two high-retrieval groups directly. Memory did not differ significantly between overt retrieval (*M* = .64, *SD* = .27) and cue-only JOLs (*M* = .65, *SD* = .23; *OR* = 0.93, *p* = .833). Transfer also did not differ following overt retrieval (*M* = .61, *SD* = .26) and cue-only JOLs (*M* = .60, *SD* = .22; *OR* = 1.03, *p* = .938). Similarly, the two low-retrieval groups did not differ significantly. Memory did not differ between the cue–target JOLs (*M* = .57, *SD* = .27) and restudy groups (*M* = .54, *SD* = .30; *OR* = 1.12, *p* = .758), nor did transfer (cue–target JOLs: *M* = .52, *SD* = .26; restudy: *M* = .50, *SD* = .27; *OR* = 1.09, *p* = .788).

Additional Comparisons with Restudy. To situate the results of our planned orthogonal contrasts within the broader FTE literature, we also report the resulting pairwise comparisons of overt retrieval and cue-only JOLs versus restudy. As a direct corollary of our preregistered hypotheses, we also tested these pairwise comparisons using one-sided tests. For memory, overt retrieval was numerically but not statistically higher than restudy (OR = 1.79, p = .054), whereas cue-only JOLs showed a significant advantage (OR = 1.93, p = .034). Similarly, both overt retrieval (OR = 1.87, p = .023) and cue-only JOLs (OR = 1.83, p = .027) produced significantly greater transfer accuracy than restudy.

Phase A Posttests

Analogous omnibus tests and contrasts were conducted for posttest accuracy for Phase A artists to provide additional context for the forward effects of interest. As these analyses were not preregistered and exploratory, all contrasts used two-tailed tests. In fact, the backward testing effect is not always observed when feedback is not provided following overt (Rowland, 2014) or covert (Yu et al., 2025) retrieval practice. We refer to posttest performance on paintings by Phase A artists as Phase A memory and Phase A transfer, not to be confused with accuracy during the interim review task in the overt retrieval condition. Descriptive statistics, full model results, and contrasts are reported in the Supplemental Materials (Table S2).

The exploratory analyses revealed no significant main effects of target condition or JOL condition, nor any interactions (all χ^2s < 1.86, all ps > .17). Unlike the forward effects observed for Phase B artists, the high- versus low-retrieval probability, overt retrieval versus restudy, and cue-only JOLs versus restudy contrasts for Phase A memory and transfer were not statistically significant (all $|zs|$ < 1.90, all ps > .06) and were numerically in the opposite direction. Nevertheless, overall posttest accuracy was descriptively higher for Phase A than Phase B posttest items, consistent with the fact that Phase A paintings were re-presented during the interim review task.

Participant-Level Indicators of Retrieval

To better understand why learning task type influenced subsequent learning, we examined self-reported retrieval use, time spent on the learning task, and their relations to posttest performance. These exploratory analyses were conducted at the participant level, with posttest accuracy expressed as proportion correct.[3] We tested key effects and interactions by comparing full models to reduced models using likelihood-ratio tests (χ^2 for ordinal logistic models of retrieval ratings, *F* tests for linear models of posttest accuracy and learning task time), and followed up with pairwise comparisons and simple effects tests using estimated marginal means. Learning task time was log-transformed for all analyses, and unless otherwise noted, time, and retrieval ratings were mean-centered when treated as predictors.

Self-Reported Retrieval Use. Mean retrieval ratings after the learning task were highest in the cue-only JOLs group (M = 7.54, SD = 1.89), followed by the restudy group (M = 7.00, SD = 2.28), and lowest in the cue–target JOLs group (M = 5.50, SD = 2.76). Ratings were significantly lower for cue–target JOLs than cue-only JOLs (−1.55, 95% CI [−2.53, −0.57], p < .001) and restudy (−1.12, 95% CI [−2.07, −0.16], p = .017), but did not differ between cue-only JOLs and restudy (p = .530). The overt retrieval group did not provide ratings because retrieval was explicitly required in that condition.

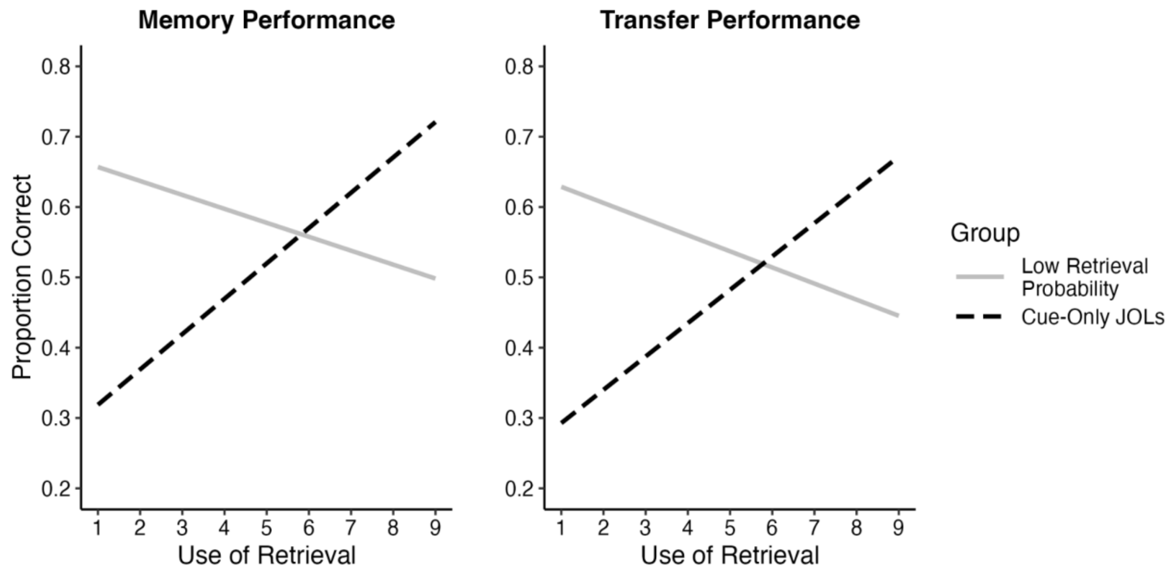
Self-reported retrieval use moderated the forward effects of learning task. Linear regression models predicting Phase B accuracy from retrieval rating, retrieval group (cue-only JOLs vs. cue–target JOLs and restudy), and their interaction revealed significant interactions for memory, $F(1,115)$ = 7.88, p = .006, and transfer, $F(1,115)$ = 9.20, p = .003.[4] As shown in Figure 4, higher retrieval ratings in the cue-only JOLs group predicted significantly better memory (b = 0.05, 95% CI [.006, .09], p = .026) and transfer (b = .05, 95% CI [.006, .09], p = .024). In contrast, retrieval were unrelated to memory (b = -0.02, 95% CI [-.04, .002], p =

[3] We aimed to keep the text focused on the main patterns of these exploratory results, so we report only the most relevant statistics in the main text. The data, analysis code, and output are available on OSF.
[4] We conducted the exploratory same moderation analyses again, separating the two low-retrieval-probability groups, which are reported in the Supplementary Materials.

.078) and negatively related to transfer (*b* = -.02 95% CI [-.04, -.002], *p* = .029) in the low-retrieval-probability groups. Thus, cue-only JOLs enhanced new learning only when participants engaged in moderate to high retrieval during the learning task.

Figure 4

Predicted Phase B Posttest Performance by Retrieval Probability Group and Self-Reported Retrieval Use in Experiment



Note. Regression lines represent estimated simple slopes from participant-level models.

Learning Task Duration. We next asked whether retrieval ratings reflected retrieval-specific processes, specifically, or more general task engagement by examining self-paced learning task duration. Overt retrieval (*M* = 1.54, *SD* = 0.49), cue-only JOLs (*M* = 1.81, *SD* = 0.93), cue–target JOLs (*M* = 1.92, *SD* = 1.02), and restudying (*M* = 1.76, *SD* = 1.29) took less than two minutes, on average, with no significant differences among conditions *F*(3,154) = 1.91, *p* = .315. Across all conditions, longer learning task time predicted better Phase B memory (*b* = .16, 95% CI [.08, .25], *p* < .001) and transfer (*b* = .15, 95% CI [.07, .23], *p* < .001), and did not vary by condition (memory: *F*(3, 150) = 0.38, *p* = .766; transfer: *F*(3, 150)

= 0.58, p = .630).

However, duration predicted retrieval ratings only when the task afforded retrieval (excluding overt retrieval participants because they did not rate retrieval) per the significant Duration × Condition interaction, $\chi^2(2) = 10.13$, p = .006. Duration strongly predicted retrieval ratings in the cue-only JOLs group (b = 1.57, 95% CI [0.73, 2.40], p < .001), but not in restudy (p = .425) or cue–target JOLs (p = .852). The retrieval rating prompt may have been interpreted differently across conditions, and are likely reflecting retrieval-specific processes for cue-only JOLs.

Joint Effects of Retrieval and Time on Task. Finally, we examined the relative contributions of retrieval use and time on task to Phase B performance within the cue-only JOLs group. When both predictors were standardized and entered simultaneously (Table 2), retrieval use remained a significant predictor of memory (β = .12, p = .025) and transfer (β = .11, p = .025), accounting for 13–14% of unique variance. In contrast, learning task duration explained less than 1% of unique variance and was no longer significant once retrieval was included. These results support that extent of retrieval, rather than general time on task, was the primary driver of forward effects in the cue-only JOL condition.

Table 2

Predicting Phase B Posttest Accuracy from Task Duration and Self-Reported Retrieval

Predictor	Estimate (β)	95% CI	*p*	Partial R^2
		Memory Performance		
log duration	.02	[-0.07, 0.11]	.603	< .01
retrieval	0.12	[0.02, 0.21]	.025	.13
		Transfer Performance		
log duration	0.01	[-0.08, 0.10]	.819	< .01
retrieval	0.11	[0.02, 0.21]	.025	.14

Note. Predictors were standardized before being entered into the model.

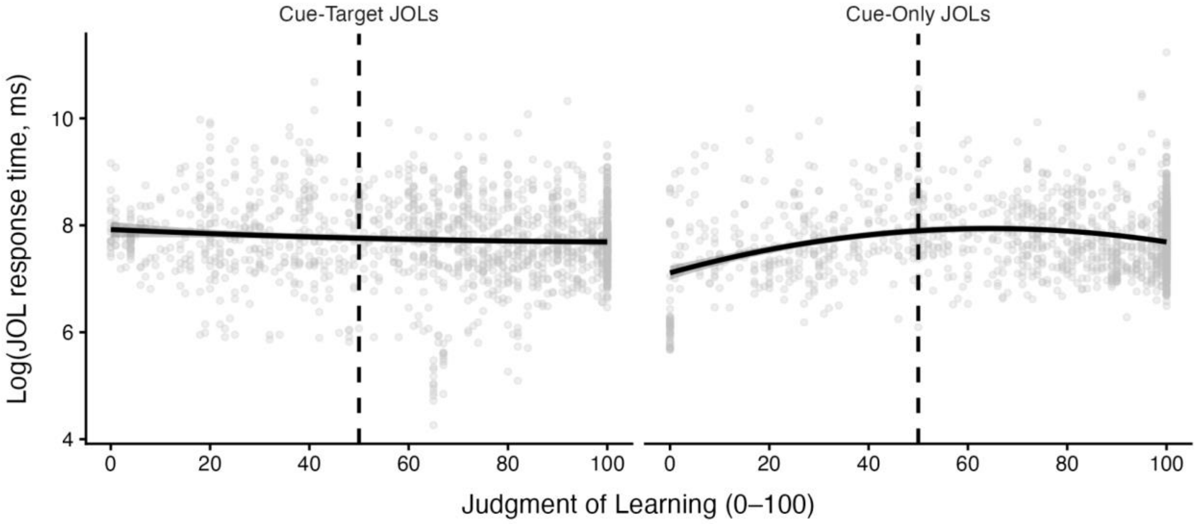
Item-Level Indicators of Retrieval

To obtain a behavioral indicator of retrieval, we additionally conducted non-preregistered analyses of the relationship between JOL magnitude and the time required to make each JOL at the item level. According to the two-stage model, cue-only JOLs involve an initial assessment of cue-familiarity or feeling of knowing the answer, and only if sufficiently high will a memory search for the target follow. The least familiar items can quickly be given a low JOL, whereas moderate JOLs reflect slower, retrieval-based processing, and if retrieval is particularly quick or fluent, a high JOL is assigned. As Son and Metcalfe (2005) demonstrated, this two-stage metacognitive process produces an inverted-U relationship between JOL magnitude and response time (RT), which is taken as a signature of retrieval-based processes (see also Metcalfe & Finn, 2008). In contrast, if cue–target JOLs function more like a restudy opportunity, then RTs should decrease monotonically as JOLs increase, reflecting longer study times for more difficult items and shorter times for easier ones. Therefore, the low JOLs were expected to differentiate cue-only and cue–target JOLs in terms of retrieval processes because higher JOLs could reflect quick retrieval or brief restudying allocated to well-learned information.

To test for overall curvature, we fit a mixed-effects model predicting log-transformed JOL RTs from JOL magnitude and its quadratic term: $log\ RT \sim JOL + JOL^2 + (1|\ Participant)$. Collapsing across JOL types, there was a significant quadratic effect, $\chi^2(1) = 34.32$, $p < .001$, establishing nonlinearity in the overall JOL–RT relationship (Figure 5). As predicted, only cue-only JOLs showed the inverted-U pattern. Low cue-only JOLs were made quickly whereas low cue–target JOLs took the longest to provide.

Figure 5

Relationship between JOL Magnitude and Response Time in Experiment 1



Note. JOL response times were recorded in milliseconds and log-transformed because of right-skew skew. Each point represents one judgment. The vertical dashed line is at the middle of the JOL scale (50). Each panel shows the fitted quadratic regression line (solid) and its 95% confidence interval (shaded) for the relationship between JOL and log-transformed response time within each condition.

The quadratic term provides only a global test of curvature but does not differentiate between the theoretically meaningful low and high JOL regions, so we used a piecewise regression to estimate separate slopes below and above a breakpoint (Son & Metcalfe, 2005). The inflection point hypothetically reflects where cue-only JOLs shift to being based on quick, initial familiarity (stage 1) to a retrieval attempt (stage 2), but its precise magnitude likely varies across tasks. Consistent with Son and Metcalfe (2005), we used the midpoint of the 0–100 JOL scale (50) as a neutral, pre-specified breakpoint.

We fit the following linear mixed-effects model: $log\ RT \sim JOL\ + Type + Segment + JOL \times Type + Segment \times Type + (1|\ Participant)$. Segment was coded as 0 for JOLs ≤ 50 and as (JOL − 50) for JOLs > 50, such that the JOL coefficient reflects the slope below the midpoint of the scale and the segment coefficient reflects the change in slope above it. Confidence intervals for the estimated slopes were obtained using parametric bootstrapping (10,000 samples) of the mixed-effects model.

JOL–RT slopes differed significantly between cue-only and cue–target JOLs below 50, $\chi^2(1) = 5.26$, $p = .022$. A 10-point increase in JOL magnitude was associated with approximately a 5% increase in RT for cue-only JOLs (95% CI [2.43, 9.63]) but a 1% decrease for cue–target JOLs (95% CI [−3.87, 2.86]). Above 50, both slopes were negative and did not differ significantly between conditions, $\chi^2(1) = 3.57$, $p = .059$. A 10-point increase in JOL magnitude was associated with approximately a 7% decrease in RT for cue-only JOLs (95% CI [−8.80, −5.06]) and a 6% decrease for cue–target JOLs (95% CI [−7.82, −3.74]).

In sum, the exploratory piecewise analyses indicate that cue-only and cue–target JOLs rely at least some distinct metacognitive processes and are consistent with the interpretation that cue-only JOLs more strongly engage covert retrieval.

Discussion

Experiment 1 supported our prediction that interim learning tasks with a high probability of engaging retrieval (overt retrieval and cue-only JOLs) enhance new inductive learning compared with low-retrieval learning tasks (cue–target JOLs and restudy), for both memory and transfer. Moreover, cue-only JOLs and overt retrieval performed similarly, as did cue–target JOLs and restudy. This pattern mirrors findings with word stems (Kubik et al., 2022) and supports the idea that cue-only JOLs benefit new learning because they elicit covert retrieval, whereas cue–target JOLs do not. Self-report data reinforce this interpretation. Participants reported significantly greater retrieval use during cue-only JOLs than during cue–target JOLs, and self-reported retrieval moderated the forward effects of cue-only JOLs. More retrieval was associated with higher memory and transfer accuracy, but only in the cue-only

JOLs condition and not the low-retrieval-probability groups. Thus, Experiment 1 provides novel, direct evidence that retrieval engagement drives the forward benefits of cue-only JOLs.

Converging evidence from item-level analyses further supports the interpretation that cue-only JOLs engaged retrieval processes. Unlike cue–target JOLs, cue-only JOLs showed the inverted-U relationship between JOL magnitude and response time, which is a signature of two-stage retrieval-based judgments (Son & Metcalfe, 2005; Metcalfe & Finn, 2008). Low JOLs were made quickly, consistent with familiarity-based judgments, whereas moderate JOLs took longer, consistent with the initiation of a memory search. High JOLs were again fast, reflecting high retrieval fluency for well-known target information. In contrast, cue–target JOLs showed the opposite pattern, with the longest latencies for low JOLs, as expected for a discrepancy-reduction approach to restudying the least well-learned information (Metcalfe & Kornell, 2005). This dissociation between JOL types provides behavioral evidence that cue-only JOLs, but not cue–target JOLs, elicit covert retrieval attempts during the learning task.

The association between retrieval ratings and Phase B posttest accuracy in the cue-only JOLs condition appears to reflect retrieval specifically rather than general task engagement. When both predictors were considered together, retrieval accounted for unique variance in posttest performance, whereas time on task did not. This aligns with research on the backward testing effect showing that it is the act of retrieval itself that enhances memory, not the time spent attempting retrieval (Vaughn et al., 2017).

One unexpected result was the relatively high retrieval ratings in the restudy group, which could be due to the phrasing of the prompt. Participants gave a rating of 9 if they “tried to remember the name of the artist from the painting”, which is clearly indicative of a retrieval attempt for cue-only JOLs but might have reflected other encoding strategies for cue–target JOLs and restudy. A restudy participant who repeated the presented artist’s name a few times before moving on to the next painting–artist pair may have interpreted their behavior as trying to remember. Indeed, time on task was positively associated with retrieval ratings for cue-only JOLs, not for cue–target JOLs or restudy.

Taken together, the results of Experiment 1 support the hypothesis that cue-only JOLs function as retrieval-based learning tasks, whereas cue–target JOLs operate more like nonretrieval tasks such as restudy. Our three planned orthogonal contrasts represent a complete test of this hypothesis because they are defined to fully account for all mean differences between conditions. Nonetheless, to situate our findings within the broader FTE literature, we also compared each retrieval-based condition to restudy individually. Both cue-only JOLs and overt retrieval significantly outperformed restudy on the Phase B posttests. The one exception was that overt retrieval did not significantly improve Phase B memory performance relative to restudy. Descriptively, however, Phase B memory accuracy was more left-skewed in the overt retrieval than the cue-only JOLs condition (Figure 3). Requiring participants to select an artist name during interim overt retrieval without providing feedback may have reinforced incorrect response (Butler & Roediger, 2008). Participants in the cue-only JOLs condition appeared to skip a retrieval attempt and provide a quick, low JOL for the paintings that felt least familiar or evoked the weakest feeling of knowing, thereby avoiding reinforcing errors (Figure 5; see also Son & Metcalfe, 2005; Metcalfe & Finn, 2008). Nevertheless, this null result should be interpreted with caution, especially because the overt retrieval and cue-only JOLs group did not differ significantly from one another on memory or transfer posttest performance.

Although Experiment 1 was not designed to adjudicate between mechanisms of the forward effect, the exploratory analyses offer some insight into how retrieval-based tasks enhanced subsequent learning of Phase B artists. One possibility is that retrieval improved learning of Phase A artists, thereby reducing proactive interference and facilitating Phase B performance (e.g., Yang et al., 2022). However, we found no evidence that the learning task influenced memory or transfer for Phase A artists (see also Ha & Lee, 2024; Lee & Ha, 2019; Yu et al., 2025), suggesting the forward benefits of high-retrieval-probability tasks observed cannot be attributed to effects in prior learning. Instead, the results point to a genuine forward effect of retrieval-based learning tasks. Experiment 2 further examined these retrieval-based explanations of the forward effect of cue-only JOLs.

Experiment 2

Experiment 2 built on Experiment 1 to test whether the forward effect of cue-only JOLs generalizes to a new domain—inductive learning of rock categories—and to examine the role of feedback. In addition to a restudy control group, participants reviewed Phase A materials with cue-only JOLs either with or without trial-by-trial correct-answer feedback. The feedback displayed the correct category name (e.g., *obsidian*) alongside the studied exemplar, equating exposure to correct answers between the JOL and restudy groups. We predicted that cue-only JOLs overall would enhance memory and transfer of new Phase B categories relative to restudy, and we further tested whether adding feedback would modulate this forward effect.

Although Experiment 1 demonstrated a forward effect of cue-only JOLs without feedback, there are strong reasons to expect that feedback could amplify this benefit, particularly given the absence of a backward testing effect from high-retrieval-probability tasks. Feedback has been shown to strengthen the backward testing effect for both overt (Rowland, 2014) and covert retrieval (Yu et al., 2025). If cue-only JOLs evoke covert retrieval, feedback could improve learning of Phase A categories, thereby reducing proactive interference (Yang et al., 2022) and sharpening attention to diagnostic features when encoding new material (Carvalho et al., 2021; Kaplan & Murphy, 2000).

On the other hand, prior work suggests that feedback is not necessary for retrieval to enhance new learning (Chan, Meissner, & Davis, 2018), which can be explained by metacognitive accounts that posit retrieval enhances new learning by altering the time, effort, and strategies participants go on to dedicate to encoding new material (e.g., Chan, Manley, et al., 2018). Specifically, the *failure-encoding-effort hypothesis* proposes that it is the experience of retrieval difficulty or failure that motivates participants to improve their subsequent study behaviors (Cho et al., 2017; for a review, see Yang et al., 2018). Feedback may therefore not substantially amplify the forward benefits of cue-only JOLs if what matters most for enhancing new learning is how learners adjust their encoding strategies in response to retrieval difficulty, rather than how much they learn from the retrieval attempts themselves.

Indeed, most studies on the FTE in inductive learning have not provided feedback during interim retrieval yet still observed robust forward effects (e.g., Yang & Shanks, 2018). The FTE may have been even larger if feedback had been provided but indirect evidence suggests this is not the case (Ha & Lee, 2024; Lee & Ahn, 2018; Lee & Ha, 2019). Little prior research on inductive learning has directly compared the forward effect of overt retrieval with and without feedback in a single experiment. One important exception is Choi and Lee (2020) who found that the forward benefit of retrieval on new inductive learning did not depend on feedback. However, neither their study nor any other experimental investigations of feedback's effect on retrieval-based new learning, to our knowledge, included cue-only JOLs as an interim review task.

Assuming cue-only JOLs enhance new learning through retrieval processes, it remains an open question whether item-by-item feedback further strengthens their forward effect. Experiment 2 directly addressed this question through two preregistered comparisons. First, if cue-only JOLs function as retrieval-based learning tasks, then JOLs (with or without feedback) should enhance Phase B memory and transfer relative to restudy. This prediction parallels Experiment 1 and serves as a replication and extension with new materials. Second, if feedback strengthens retrieval-based processes by improving memory for Phase A categories, then cue-only JOLs with feedback should improve new learning compared to cue-only JOLs without feedback. Conversely, if the forward effects of cue-only JOLs arise primarily from metacognitively guided changes in encoding prompted by experiencing retrieval difficulty, feedback should confer no additional benefit. Together, these comparisons allow us to test not only whether cue-only JOLs reliably produce a forward effect on inductive learning, but also whether feedback modulates that effect, thereby clarifying the mechanisms through which JOL-based learning operates.

The preregistration is available at the Open Science Framework: https://osf.io/7hufq/?view_only=5da2bb01a64743acbf8cb1d136cb67d7, as are the materials, data, and analysis scripts [anonymized for peer review, requiring separate links]: https://osf.io/ka4nu/?view_only=4179aadc0239419482ab2f0979926d80.

Method

Experiment 2 introduced several key changes relative to Experiment 1. We replaced paintings with photographs of rocks to test the generality of cue-only JOLs across domains, and we narrowed the design to focus on restudy and cue-only JOLs (with or without feedback), omitting the cue–target JOLs and overt retrieval groups. To probe potential metacognitive consequences of different learning tasks, participants reported at the end of Phase A their planned effort and motivation for Phase B learning, as well as the percentage of Phase B rock photos they believed they would successfully learn (an ease-of-learning [EOL] judgment). To capture possible downstream changes in encoding effort or strategy, Phase B study trials were self-paced rather than fixed. Finally, we removed the retrieval-use question administered after Phase A because it was difficult to phrase comparably across conditions and might itself influence how participants approached Phase B. In particular, asking about retrieval could have disproportionately prompted restudy participants to shift their study approach toward memorizing exemplars for later recall.

Participants

As in Experiment 1, participants were recruited through Moodle, Unipark, and SONA. Participants were students from the FernUniversität in Hagen, the University of Würzburg, and the University of Bielefeld. Using the same criteria as Experiment 1 participants were excluded for previous participation (*n* = 1), using external resources (*n* = 4), or having extremely low memory and transfer posttest performance alongside a lack of focus (*n* = 3), unserious participation (*n* = 4), or speaking German less than 10 years (*n* = 0). Among the 178 participants recruited, 166 met the inclusion criteria (128 female, 35 male, 2 gender-diverse, 1 no answer; age: *M* = 28.04 years, *SD* = 9.81, range = 18 to 61 years). They were fluent in German and received course credit for participating.

Design

The experiment used a one-factorial, between-subjects design. Participants were randomly assigned to one of three interim learning task conditions in Study Phase A: cue-only JOLs with feedback, cue-only JOLs without feedback, or restudy. The primary dependent variable was item-level classification accuracy on the transfer and memory posttests for paintings by Phase B rock photos. Time on the interim learning task, planned effort and motivation for Study Phase B, and time spent studying Phase B rocks were examined as potential moderators.

Materials and Procedure

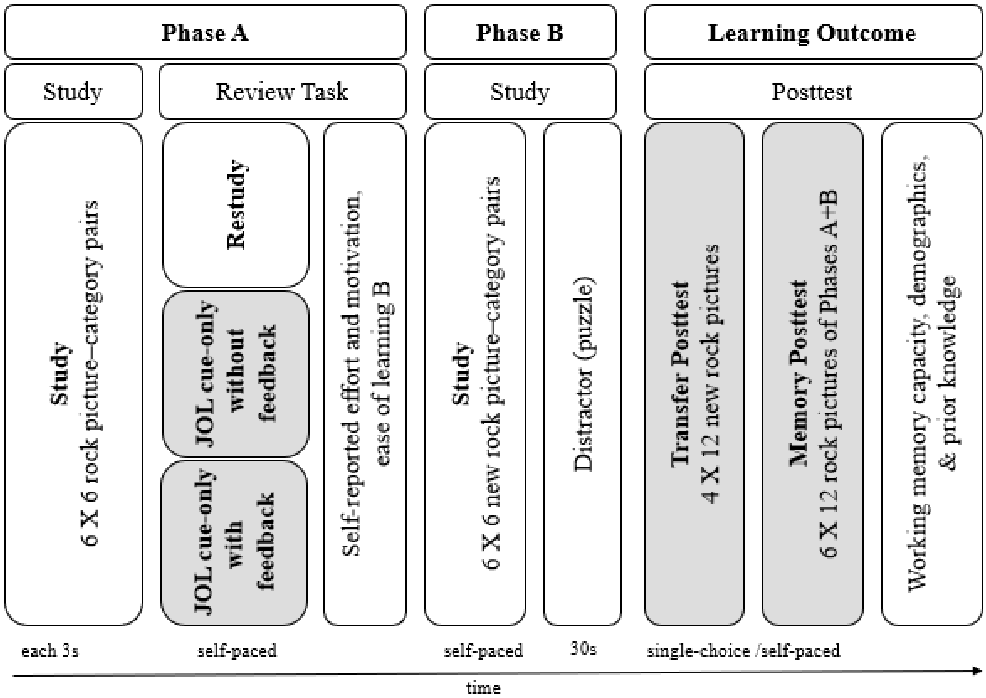
The procedure (Figure 6) was highly similar to Experiment 1, but the study materials were 120 pictures of different types of rocks (10 by each of 12 rock categories) from Kang and colleagues (2023).[5] In Study Phase A, participants viewed 36 rock photos (six per each of

[5] The rock pictures used in our study and Kang et al. (2023) were adapted from Miyatsu et al. (2019) (retrieved from https://osf.io/9vg8m/). Nosofsky et al. (2018) originally created these pictures.

six categories), each labeled by the name of the rock category to which it belonged. The rock photo–category pairs were presented individually, in a random order, in the center of the computer screen, for 3 seconds, with a 0.5-second blank screen between pairs.

Figure 6

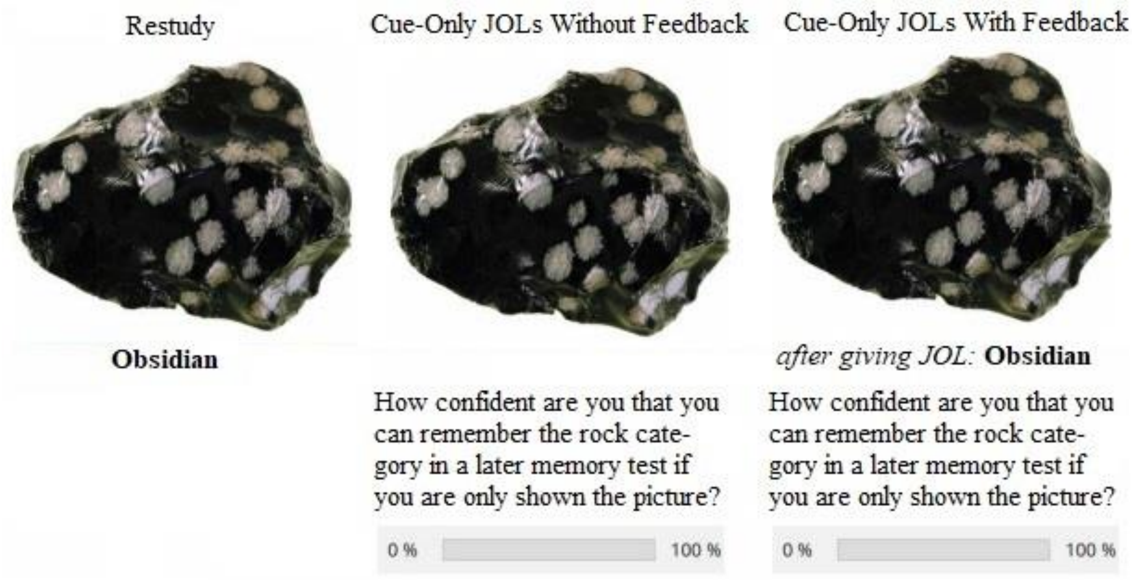
Experiment 2 Procedure



Next, participants completed one of the three interim learning tasks, reviewing the same 36 rocks from Study Phase A in a new random order (Figure 7). *Restudying* involved viewing each rock–category pair again without responding. For *cue-only JOLs without feedback*, only the rock photo was shown and rated the likelihood (0–100%) that they would remember the rock category name if presented with the rock photo on a later memory posttest. Restudying and JOLs were self-paced. The procedure was the same for *cue-only JOLs with feedback* except that after each self-paced JOL, the rock picture remained on the screen while the JOL prompt disappeared and the name of the correct rock category was displayed in black letters for a fixed two seconds. Participants could not change their JOLs during this time. A 0.5-second blank screen between all trials.

Figure 7

Examples of the Three Learning Tasks in Experiment 2



After completing the review phase, participants rated planned effort, motivation, and provided an EOL judgment. They then studied 36 new rock photos (six by each of six new categories) using the same procedure as in Phase A before a 30-second distractor task (nine-piece digital puzzles). Participants first completed the transfer posttest, classifying 48 new rocks (four per category) presented in a random order by selecting the correct category name from 12 options. The memory posttest on the 72 previously studied rock photos followed using the same procedure. Finally, participants answered demographic and data quality check questions.

Instruments and Measures

Learning outcomes were assessed on the transfer posttest of 48 unseen rock photos and the memory posttest of the 72 rocks shown during Phases A and B. On each trial, participants selected the correct rock category from a list of 12 options. Accuracy was coded at the item level (1 = correct, 0 = incorrect), which served as the dependent variable in the primary mixed-effects analyses. Accuracy was also summarized as the proportion of correct responses per participant for descriptive statistics and additional exploratory analyses.

At the end of Study Phase A, participants rated how much effort they planned to invest in the next block studying new rock photos and categories from 1 (“very low”) to 7 (“very much”) and how motivated they were to do so (1 = “not motivated at all” to 7 = “very motivated”). The ease-of-learning judgment (EOL) asked participants what percentage of the upcoming new rock photos (0–100%) they believed they would correctly identify on the posttest.

Learning task duration was measured as the total time spent on the self-paced cue-only JOLs or restudy, recorded in milliseconds and converted to minutes. Phase B study duration was measured as the total time spent on the self-paced initial study trials in Phase B, recorded in milliseconds and converted to minutes.

Preregistered Analysis Plan

The preregistered analyses closely followed that of Experiment 1. Again, our primary analyses focused on memory and transfer posttest accuracy for Phase B rock types. Each model was specified as $Accuracy \sim Condition + (1| Participant) + (1 \mid Stimulus)$, where *Condition* was a three-level factor. Planned orthogonal contrasts were defined *a priori* to test preregistered hypotheses. Our central goal was to evaluate our hypothesis that cue-only JOLs function as retrieval-based learning tasks, and test whether their forward benefit depends on feedback. Therefore, the first preregistered contrast tested a directional hypothesis for *review type*, comparing the JOL groups (with or without feedback) to the restudy group with weights of +0.5, +0.5, and −0.5, respectively.[2] An additional preregistered contrast directly compared cue-only JOLs with and without feedback (+1, −1, 0).

To clarify specific effects raised during peer review, we additionally compared cue-only JOLs with feedback (+1, 0,−1) and cue-only JOLs without feedback versus restudy (0, +1, −1). Although these pairwise contrasts are not independent of the planned JOLs (with or without feedback) versus restudy contrast, they are derived from our preregistered predictions.

If the JOLs conditions overall outperform restudy but do not differ from each other, then each JOL condition should individually exceed restudy. Accordingly, we used one-sided tests for the JOLs versus restudy contrast, and for pairwise comparisons of each JOLs condition versus restudy. All other contrasts were two-sided and reported with 95% confidence intervals. Additional exploratory analyses are described in the Results section and follow the same approach as Experiment 1 unless otherwise noted.

Results

Judgments of Learning

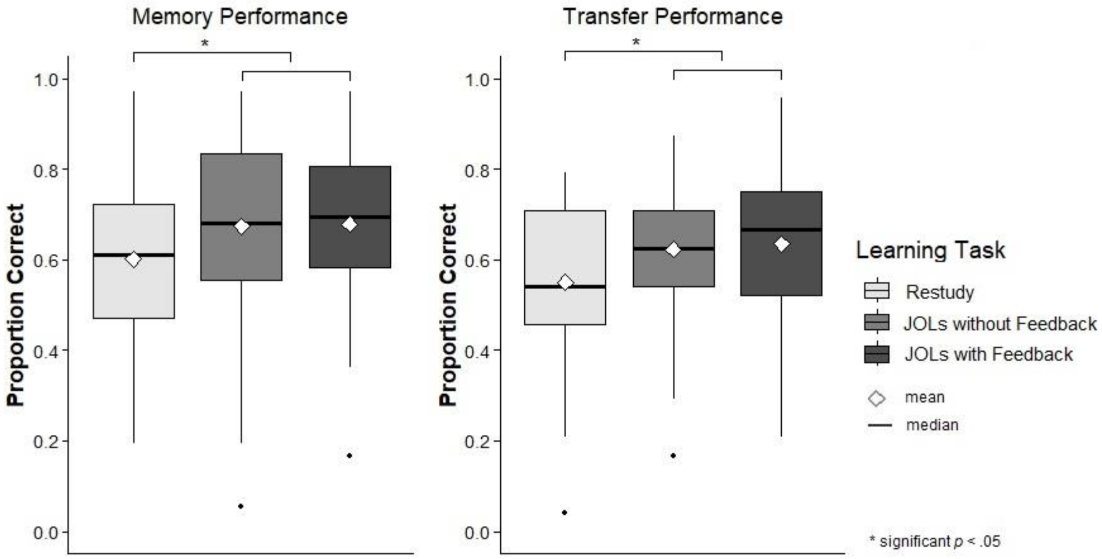
We report descriptive statistics about judgments of learning for completeness. Participants in the cue-only JOLs groups provided similar average JOLs during the learning task regardless of whether feedback was provided after each JOL (M = 79.50%, SD = 15.10%) or not (M = 76.96%, SD = 16.73%). After the interim review task, all participants also provided a single EOL, judging the percentage of the upcoming Phase B rocks they would be able to correctly label on the memory posttest. EOLs were similar following restudy (M = 54.82%, SD = 21.48%), cue-only JOLs without feedback (M = 51.55%, SD = 20.39%), and cue-only JOLs with feedback (M = 57.25%, SD = 18.79%).

Phase B Posttests

Our primary analyses focused on posttest accuracy for Phase B rock categories (Figure 8). Overall, there was a statistically significant main effect of condition on transfer, $\chi^2(2)$ = 6.27, p = .04, but not memory, $\chi^2(2)$ = 4.83, p = .089. Full model results and fit statistics are reported in the Supplemental Materials (Table S3). Planned contrast analyses tested our preregistered, theory-driven predictions about specific combinations of learning task conditions (Table 3).

Figure 8

Memory and Transfer Posttest Accuracy in Experiment 2



Note. Proportion correct was aggregated by participant for the boxplots.

Cue-only JOLs versus Restudy. As predicted, there were significant forward benefits of cue-only JOLs compared to restudy. The one-sided test of the planned contrast revealed that cue-only JOLs (with and without feedback: *M* = .68, *SD* = .18) enhanced Phase B memory compared to restudy (*M* = .60, *SD* = .19; *OR* = 1.47 *p* = .017). Together, cue-only JOLs (*M* = .63, *SD* = .16) also outperformed restudy on Phase B transfer (*M* = .55, *SD* = .17; *OR* = 1.43, *p* = .009).

Table 3

Experiment 2 Contrasts of Posttest Accuracy for Phase B Rocks

Contrast	Odds Ratio	95% CI (two-sided) / LB_{90}% (one-sided)	*z*	*p*
		Memory Performance		
Cue-only JOLs vs. restudy*	1.47	> 1.17	2.12	.017
Feedback vs. no feedback*	1.15	[0.75, 1.75]	0.65	.518
No feedback vs. restudy	1.37	> 1.05	1.53	.063
Feedback vs. restudy	1.58	> 1.20	2.12	.017
		Transfer Performance		
Cue-only JOLs vs. restudy*	1.43	> 1.18	2.36	.009
Feedback vs. no feedback*	1.18	[0.83, 1.68]	0.94	.348
No feedback vs. restudy	1.32	> 1.05	1.58	.057
Feedback vs. restudy	1.56	> 1.24	2.47	.007

Note. Wald *z*-tests from generalized linear mixed-effects models with random intercepts for participants and stimuli. Cue-only JOLs refers to the two JOL groups combined. Feedback and No feedback refer to the cue-only JOLs groups with and without feedback, respectively. *Indicates preregistered contrasts.

Feedback Effects. Cue-only JOLs with versus without feedback resulted in similar new learning (Figure 8). Memory did not differ significantly between cue-only JOLs with feedback (M = .68, SD = .17) and without feedback (M = .67, SD = .19; OR = 1.15, p = .518). There was no significant difference in transfer following cue-only JOLs with feedback (M = .64, SD = .16) and cue-only JOLs without feedback (M = .62, SD = .16; OR = 1.18, p = .348).

Additional Comparisons with Restudy. To situate the results of our planned orthogonal contrasts within the broader FTE and JOL reactivity literatures, we also report the resulting pairwise comparisons of each cue-only JOLs condition vs. restudy. As a direct corollary of our preregistered hypotheses, we also tested these pairwise comparisons using one-sided tests. Cue-only JOLs with feedback significantly enhanced Phase B memory (OR =

1.58, p = .017) and transfer (OR = 1.56, p = .007) compared to restudy. However, cue-only JOLs without feedback only demonstrated numerical, but not statistically significant, advantages over restudy for memory (OR = 1.37, p = .063) and transfer (OR = 1.32, p = .057).

Phase A Posttests

As in Experiment 1, we conducted orthogonal contrasts of posttest performance for Phase A rock categories, which we refer to as Phase A memory and Phase A transfer. The design of Experiment 2 was built on the finding that feedback amplifies the backward testing effect of overt retrieval (Rowland, 2014) and covert retrieval (Yu et al., 2025), which may not be observed in the absence of feedback. Therefore, we hypothesized that feedback would also amplify the backward effects of cue-only JOLs, which may not improve Phase A memory without feedback. Accordingly, we report separate one-sided tests of contrasts of cue-only JOLs with feedback versus cue-only JOLs without feedback and versus restudy. A two-sided test was used for the contrast of cue-only JOLs without feedback versus restudy. Descriptive statistics, full model results, and contrasts are reported in the Supplemental Materials (Table S4).

There was no significant overall effect of learning task on Phase A memory, $\chi^2(2)$ = 3.74, p = .154, or transfer $\chi^2(2)$ = 3.55, p = .169. As expected, cue-only JOLs with feedback significantly improved memory (OR = 1.35, p = .037) and transfer (OR = 1.30, p = .038) compared to cue-only JOLs without feedback. However, cue-only JOLs with feedback did not significantly outperform restudy on the Phase A memory or transfer posttests and there were no differences between cue-only JOLs without feedback and restudy, either (all ps > .134).

Time on Task

Similar to Experiment 1, we explored whether time-on-task differed by learning task condition or predicted new learning, and if so, whether the time–learning association was moderated by condition.[3] Times were log-transformed for all analyses, and unless otherwise noted, time, effort, and motivation were mean-centered when treated as predictors.

Learning Task Time. Participants spent on average less than two minutes on restudy ($M = 1.67$, $SD = 1.37$) and on cue-only JOLs without feedback ($M = 1.87$, $SD = 0.74$). In the cue-only JOLs with feedback group, the average time spent on the learning task was 2.26 minutes ($SD = 1.29$), excluding the two seconds of feedback after each of the 36 items. Unlike Experiment 1, there was a significant overall effect of condition on learning task duration, $F(2, 163) = 10.27$, $p < .001$, with participants spending significantly more time providing JOLs than restudying ($p < .001$).

Phase B Study Time. The average time to study the 36 new pictures was 2.15 minutes ($SD = 1.88$) in the restudy group, 1.90 minutes ($SD = 1.78$) in the cue-only JOLs without feedback group, and 1.90 minutes ($SD = 1.17$) in the cue-only JOLs with feedback group. Study duration did not differ significantly across conditions, $F(2, 163) = 0.25$, $p = 0.777$.

We examined the relative contributions of learning task time and Phase B study time on new learning, after controlling for learning task (JOLs vs. restudy). When both standardized time predictors were entered into the same linear regression model along with learning task, only Phase B study time remained a significant predictor of Phase B memory ($\beta = .06$, $p < .001$, partial $R^2 = .071$), whereas learning task duration did not ($R^2 < .01$). A similar pattern emerged for transfer (Phase B study time: $\beta = .04$, $p = .013$, partial $R^2 = .04$; learning task duration: $R^2 < .01$). Learning in Phase B was perhaps unsurprisingly driven more by time spent studying Phase B material than reviewing Phase A material through JOLs or restudy.

Effort and Motivation

Experiment 2 also explored self-report measures of metacognitive processes that could be potentially affected by the different learning tasks. Separately, both effort ($p = .042$) and motivation ratings ($p = .014$) significantly positively predicted Phase B study times, supporting their use as indicators of engagement. Effort ratings differed across the three learning tasks ($\chi^2(2) = 10.28$, $p = .006$), with significantly lower planned effort following restudy than cue-only JOLs without feedback (−1.11, 95% CI [−1.94, −0.29], $p = .004$). However, effort did not differ between cue-only JOLs with feedback ($M = 5.59$, $SD = 1.13$) and restudy ($M = 5.02$, $SD = 1.53$, $p = .100$) or between cue-only JOLs with and without feedback ($M = 5.76$, $SD = 1.33$, $p = .463$).

On its own, greater planned effort predicted significantly better Phase B memory ($b = 0.02$, 95% CI [.002, .04], $p = .027$), but not transfer ($b = 0.02$, 95% CI [0, .04], $p = .059$). Learning task did not moderate the effort–posttest performance association (memory: $F(1, 162)$ 0.12, $p = .729$; transfer: $F(1, 162) = 0.03$, $p = .874$), and effort no longer predicted memory or transfer after controlling for learning task (memory: $p = .084$, transfer: $p = .200$).

Motivation did not differ significantly by learning task condition, $\chi^2(2) = 0.38$, $p = .828$. Motivation also did not predict Phase B memory performance on its own (bs < .01, ps > .198), and the motivation–posttest performance relationship was not moderated by learning task (Fs < 1.26, ps > .263). Thus, while cue-only JOLs increased learners' planned effort, effort and motivation did uniquely explain new learning beyond the influence of the learning task itself.

Item-Level Indicators of Retrieval

As in Experiment 1, we explored the item-level relationship between the magnitude and response times of JOLs. We analyzed only data from the no-feedback condition because prior research does not specify how feedback might alter the threshold for initiating retrieval or the duration of memory search. Including the feedback condition would not have meaningfully informed interpretation of the cue-only JOLs without feedback, for which we had specific, theory-driven predictions.

We again tested for overall curvature with a mixed-effects model predicting log-transformed RTs from JOL magnitude and its quadratic term, which yielded a significant quadratic effect, $\chi^2(1) = 45.89$, $p < .001$. We then fit a piecewise regression model, $log\ RT \sim JOL\ + Segment + (1|\ Participant)$, to separately estimate the slopes below and above the midpoint of the JOL scale (50). The slope was significantly positive 50, $\chi^2(1) = 5.75$, $p = .016$, and significantly negative above 50, $\chi^2(1) = 149.78$, $p < .001$. A 10-point increase in JOL magnitude was associated with an approximately 4% increase in JOLs below 50 (95% CI [0.75, 7.15]) but a 9% decrease in JOLs above 50 (95% CI [−9.87, −7.16]). Together, these results indicate a clear inverted-U relationship between JOL magnitude and response time, consistent with the two-stage model of JOLs.

Discussion

Experiment 2 replicated and extended the forward benefit of cue-only JOLs relative to restudy with new materials. As in Experiment 1, these forward effects could not be attributed to differences in time spent on the learning task pointing instead to differences in the cognitive processes engaged during that time. These findings align with our hypothesis that retrieval-based mechanisms underlie the forward benefit of cue-only JOLs, which appear to engage cognitive processes similar to overt retrieval (Dougherty et al., 2018; Kimball & Metcalfe, 2003). Specifically, analysis of the no-feedback condition again revealed an inverted-U relationship between JOL magnitude and response time. This is consistent with the two-stage model of JOLs with an initial familiarity-based judgment that is only followed-up by a memory search if the material evokes sufficient familiarity or feelings of knowing the answer (Son & Metcalfe, 2005). When retrieval is initiated, longer, and more effortful searches tend to produce moderate JOLs, compared to fluent, easily accessed targets that yield higher JOLs and faster judgments.

Our second main question concerned whether providing item-by-item feedback would amplify the forward effect of cue-only JOLs. It did not, which mirrors Choi and Lee’s (2020) finding that feedback did not moderate the FTE in inductive learning and suggests that feedback is not a necessary condition for cue-only JOLs to benefit new learning. One possible explanation is that feedback reduced retrieval effort during JOLs by allowing participants to wait for forthcoming answers, but this explanation is unlikely because JOL participants who received feedback actually spent numerically *more* time on the task than those without feedback.

From the perspective of interference-reduction accounts (Yang et al., 2022), feedback might have improved learning of the Phase A categories, thereby reducing proactive interference and facilitating new encoding. Indeed, Phase A memory and transfer were somewhat higher in the feedback condition than in the no-feedback condition, yet this did not translate into larger forward effects. Moreover, cue-only JOLs with feedback and restudy did not differ in Phase A performance, yet the feedback condition still showed a forward benefit relative to restudy. These results suggest that the forward benefits of cue-only JOLs are not mediated by improved retention of prior material.

Metacognitive accounts of the FTE can provide a better explanation for the forward JOLs effect we observed. For instance, according to the failure-encoding-effort hypothesis, retrieval difficulty can prompt changes in subsequent encoding effort and strategies (Cho et al., 2017; Yang et al., 2018). Feedback would not be essential because the key driver of new learning lies in the retrieval experience, not its direct backward benefit to memory. Consistent with a metacognitive explanation, JOL participants reported higher planned effort for Phase B study than restudy participants, and these effort ratings predicted subsequent study times at the individual level. However, Phase B study times did not differ significantly between learning task conditions, and planned effort no longer predicted new learning once learning task (JOLs vs. restudy) was accounted for. This suggests that effort ratings can indicate metacognitive processes that change with the learning task, but effort is not the direct mechanism underlying new learning.

The forward benefits of cue-only JOLs may thus reflect qualitative changes in subsequent study strategies rather than just study time (Chan, Manley, et al., 2018; Finley & Benjamin, 2012; Yang et al., 2022). Learners differ in their tendency to memorize exemplars versus abstract underlying rules, which affects memory and transfer (McDaniel et al., 2022). People can adapt their approach with task experience (Gouravajhala et al. 2020), and the retrieval processes we hypothesize are involved in cue-only JOLs may prompt such strategy shifts. Future research should examine whether planned effort or other metacognitive measures predict qualitative changes in study behavior, even when dissociated from total study time.

In sum, Experiment 2 strengthens the conclusion that cue-only JOLs function as retrieval-based learning tasks that facilitate new inductive learning relative to restudy. The absence of any feedback effect, despite modest backward benefits, argues against interference-reduction accounts and points instead to metacognitive mechanisms as the primary driver of these forward effects.

General Discussion

The present research clarifies when and why judgments of learning (JOLs) can enhance new learning. Previous JOL reactivity research has demonstrated that making JOLs can improve memory for previously studied materials, although their memorial benefits are not universal (Double et al., 2018; Ingendahl et al., 2025; Zaho et al., 2022). Far less is known about their effects on subsequent learning and the existing studies have produced mixed results. In particular, research of the forward effects of JOLs on inductive category learning (Ha & Lee, 2024; Lee & Ha, 2019; Wang et al., 2023) has varied in the JOL grain sizes used (item-level, category-level, and aggregate), the completeness of information provided for the judgment (cue-only vs. cue–target), and how new learning was assessed (memory vs.

transfer). Because these factors were not systematically manipulated within a single design, it remains unclear when JOLs facilitate new learning, and why their effects sometimes fail to emerge.

We tested the retrieval-based mechanisms proposed to underlie some types of JOLs by holding judgment grain size constant at the item level and systematically varying retrieval demands with cue-only and cue–target formats, alongside restudy and overt retrieval for comparison. We also assessed both memory and transfer to evaluate whether JOL-induced retrieval selectively benefits item-specific memory versus generalized category learning. Across two experiments, we found that cue-only JOLs produced robust forward benefits on new inductive learning, replicating and extending the effects previously observed with verbal materials (Kubik et al., 2022). In Experiment 1, cue-only JOLs enhanced both memory and transfer to the same degree as overt retrieval, whereas cue–target JOLs yielded outcomes comparable to restudy. In Experiment 2, cue-only JOLs again outperformed restudy, and this benefit was unchanged when feedback followed each judgment.

Importantly, patterns in JOL response times were indicative of retrieval processes during cue-only JOLs in both experiments. In Experiment 1, cue–target JOLs response times did not suggest retrieval, which was reflected in lower self-reported retrieval ratings in the cue–target than cue-only condition. Finally, greater self-reported retrieval during cue-only JOLs predicted new learning above and beyond overall time on task. These findings demonstrate that not all metacognitive judgments are created equal: Item-specific JOLs promoted new inductive learning most reliably when they engaged retrieval-like processing, though other mechanisms may also contribute.

A focus on retrieval-based processes helps explain inconsistencies in prior findings on the forward effects of JOLs in inductive learning. Studies that included complete cue–target information during JOLs have found no advantage over restudy (Ha & Lee, 2024; Lee & Ha, 2019), whereas Wang et al. (2023, Experiment 3) reported forward benefits when JOLs were made from cues alone. This pattern aligns with our findings and supports the view that cue–target JOLs fail to elicit the covert retrieval processes that promote new learning. A seemingly contradictory result comes from Ha and Lee (2024), who found no forward benefit of cue-only JOLs unless they were paired with overt retrieval. Notably, Wang and colleagues (2023) used a memory posttest, whereas Ha and Lee (2024) used a transfer posttest and found no forward benefit, raising the possibility that cue-only JOLs emphasize item-specific processing (e.g., memorizing visual features) at the expense of interitem relational processing that support generalization (Double et al., 2025). However, because our experiments included both memory and transfer assessments and revealed cue-only advantages for each, this explanation appears unlikely.

Instead, the divergent results across studies likely reflect differences in the metacognitive processes engaged by how the JOL prompt was phrased (e.g., Ozuru et al., 2012). In our experiments, participants rated how likely they were to *remember* the category name when shown a studied exemplar (see also Wang et al., 2023). This “remember” phrasing likely encouraged active retrieval of the missing category name with cue-only JOLs. In contrast, Ha and Lee (2024) asked participants to rate how likely they were to *identify* new exemplars of the same category, which may have fostered familiarity-based or fluency-based judgments rather than retrieval (Besken, 2016; Fazendeiro et al., 2005). As our Experiment 1 showed, the degree of retrieval during cue-only JOLs strongly predicted new learning. From this perspective, Ha and Lee’s (2024) seemingly discrepant findings reflect the same underlying principle: Retrieval engagement, rather than judgment making per se, is the critical mechanism driving the forward effects of JOLs (Dougherty et al., 2005).

Individual Differences in Self-Reported Retrieval

Although cue-only JOLs generally prompted retrieval-like processing, participants varied considerably in the degree to which they reported attempting retrieval, which moderated the effectiveness of cue-only JOLs. Therefore, the forward benefits of metacognitive judgments depend not just on the judgment prompt but on how learners approach them. This finding is consistent with self-regulated learning theories that emphasize the interaction between metacognitive awareness and strategy use in optimizing learning outcomes (e.g., Zimmerman, 2002).

It remains an open question why participants varied in the degree to which they engaged in retrieval when providing cue-only JOLs. Retrieval is generally considered effortful, and more so than restudying (Hui et al., 2022), so motivation may explain some of the variability (Carpenter et al., 2022). Alternatively, some participants might have been sufficiently motivated to thoroughly attempt to remember the artist name or rock type if directly prompted to do so, but they do not self-initiate retrieval during cue-only JOLs because they are not aware of its benefits (Carpenter et al., 2022). Exploring how individual differences in motivation and self-regulated learning skills with the forward effects of JOLs could inform these theoretical questions as well as personalized educational interventions involving metacognitive reflection.

Mechanisms of Retrieval-Based Learning from JOLs

Drawing on the FTE literature, there are myriad explanations for why the retrieval-based processes associated with cue-only JOLs would enhance new learning (for reviews of FTE theories, see Chan, Meissner & Davis, 2018; Yang et al., 2018). Forward effects of cue-only JOLs have been attributed to mechanisms relating to memory for the Phase A categories. Such a memory-based, interference account suggests that retrieval improves memory for the tested information, thereby reducing proactive interference and/or facilitating integration with new, subsequently-presented material (e.g., Yang et al., 2022). This explanation was not supported in our study as we did not find any evidence for a backward effect with natural categories on the posttest in Experiments 1 and 2. Note, however, that because we did not include a condition that only completed Study Phase A, we cannot cleanly determine whether cue-only JOLs improved learning of the Study Phase A categories compared to cue–target JOLs or restudy. Additional research is needed to extend this memory-based account of the FTE to forward effects of JOLs.

An alternative, metacognitive explanation of the FTE is more consistent with our results. It suggests that interim covert retrieval elicited by cue-only JOLs enhances new learning indirectly by motivating learners to pay more attention, invest more effort, or use more effective encoding strategies during subsequent learning blocks (e.g., Cho et al., 2017). Specifically, the effort associated with attempting to covertly retrieve the target information for cue-only JOLs, and perhaps failing to do so, reveals to learners limitations in their memory and thus prior encoding. Consistent with this account, Experiment 2 participants planned to put in more effort studying new Phase B material following cue-only JOLs than restudy. Miller and Geraci (2014) found that even just one experience of retrieval failure is sufficient to eliminate overconfidence in memory performance for an entire task, not just the failed item. Therefore, even though participants likely experienced only a limited amount of retrieval failure (e.g., overt retrieval accuracy was 79% and cue-only JOLs were 72% in Exp. 1), it could be sufficient to prompt changes in subsequent studying effort and strategies. Previous research has found that the forward benefits of cue-only JOLs can emerge with one category-level JOL per category or even just one global JOL (Lee & Ha, 2019), although additional research is needed to clarify how the forward effects of JOLs depend on the amount of retrieval attempts versus the subjective experience of difficulty.

Finally, other retrieval-based mechanisms beyond metacognition could also contribute to the forward effects of cue-only JOLs in the present research, such as the restoration of encoding resources before new material is introduced (Pastötter et al., 2011). Future research should examine whether resource-based accounts of the forward testing effect (Chan, Meissner, & Davis, 2018) can also explain the forward benefits of cue-only JOLs.

Limitations and Future Directions

Experiment 1 found no significant differences in the forward effects of cue-only JOLs and overt retrieval. We designed the overt retrieval task such that participants chose the artist's name from a list to keep the procedure as similar as possible to the cue-only JOL task (i.e., clicking the screen to provide a response). The recognition format of overt retrieval may have attenuated its effect on new learning, since free and cued recall tests generally strengthen memory more than recognition tests (Rowland, 2014). Nevertheless, recognition tests produce a robust backward testing effect (Rowland, 2014), and the FTE has been observed with natural category learning when the learning task is a recognition test (Wang et al., 2023; Yang & Shanks, 2018). Future research should investigate the potential moderating role of learning task format and difficulty (e.g., Hausman et al., 2025).

Even if the forward effects of overt retrieval could be larger with a different test format, the self-report retrieval question supports our interpretation that cue-only JOLs also promote new learning through retrieval. However, one limitation of the present study is that the retrieval rating question asked participants the degree to which they "tried to remember the name of the artist from the painting". It is clearly indicative of a retrieval attempt for cue-only JOLs but might have reflected effort invested in other encoding strategies in the cue–target JOLs and restudy conditions. Indeed, self-reported retrieval did not differ between restudy and cue-only JOLs. Furthermore, the correlation between total time on the learning task and retrieval rating was positive, negative, and near zero in the cue-only, cue–target, and restudy conditions, respectively, suggesting that participants did not interpret the question the same way across conditions. Other types of evidence for retrieval during JOLs beyond self-reports is needed. Future research should use process-tracking methods, such as eye-tracking or think-aloud protocols, to gather additional converging evidence for how learners interact with the material during the JOL task, and how these interactions relate to future learning outcomes. Think-aloud protocols have been successfully used to identify how different learning tasks affect the cues the people use to make their metacognitive judgments (Thiede et al., 2010).

The same process-tracking methods could also be used to understand the effort, attention, and strategies participants use to encode the Study Phase B exemplars. Although Experiment 2 found no evidence that cue-only JOLs led to greater subsequent study effort compared to restudying in terms of Study Phase B encoding times, it remains possible that there were qualitative differences in encoding strategies that we did not capture. For example, eye-tracking during Study Phase B could be a useful tool for determining whether different interim learning tasks affect the degree to which participants pay attention to features of the exemplars that discriminate between different categories (Carvalho & Goldstone, 2017).

Practical Applications

Given growing evidence that metacognitive judgments can directly and indirectly improve learning and memory, digital educational tools such as adaptive learning platforms and intelligent tutoring systems are incorporating frequent metacognitive prompts (Klar et al., 2024). For instance, InQuizitive is a commonly-used homework and assessment platform in post-secondary education and essentially uses cue-only JOLs: For each question, students must rate their level of understanding before they can submit their answer and receive feedback. One of the key implications of our findings is that cue-only judgments themselves can act as effective interim retrieval opportunities that enhance subsequent learning. To capitalize on these forward benefits, educators and designers might pair these assignments with follow-up learning tasks (e.g., a brief reading, a short video, a few motivating examples) to introduce new course content.

However, we also observed that not all participants reported engaging in high levels of retrieval during cue-only JOLs, and quick, low JOLs likely reflected skipped retrieval attempts. Consistent with the two-stage model of JOLs, metacognitive judgements do not necessarily enhance future learning, specifically in more complex materials such as multiple text learning in which memory retrieval is likely truncated when not explicitly prompted (Hausman & Kubik, 2023). Thus, digital tools could flag rapid low-confidence responses and nudge learners to attempt retrieval before viewing feedback. This could be especially useful in low-stakes, completion-based assignments where motivation to generate an answer before submitting it could be lower.

Finally, even if students are not overtly instructed to provide item-by-item judgments JOLs in lectures or assignments, learners often engage in spontaneous or implicit metacognitive monitoring without instruction (e.g., Dunlosky & Metcalfe, 2009; Koriat, 1997). For example, when using flashcards or study guides, students view questions, terms, or prompts, and automatically assess their knowledge of the target information, effectively performing cue-only JOLs without being explicitly prompted. The results of Experiment 1 suggest that, unlike cue-only JOLs, cue–target JOLs are no better for new learning than restudy. Learning materials should delay exposure to the answer to increase the chance that students' implicit, automatic metacognitive judgments are based on retrieval. Together, these applications illustrate how understanding the retrieval-based mechanisms of metacognitive judgments can inform relatively simple design choices for learning tools and materials that more effectively promote lasting, transferable learning.

Conclusion

Reflecting on one's learning is not a neutral process; providing judgments of learning can directly affect learning. The present study advances our understanding of the forward effects of cue-only JOLs, particularly on new inductive learning, and highlights the importance of retrieval-based strategies for effective learning. These findings add to the growing body of evidence supporting the forward benefits of JOLs and their applicability to diverse materials, while also raising questions about the role of feedback in modulating learning from JOLs. By addressing these questions, future studies can further elucidate the mechanisms underlying JOL-based learning and expand the practical applications of this metacognitive strategy, ultimately bridging the gap between metacognitive monitoring, active learning strategies, and long-term transferable learning.

Declarations

The study was conducted in accordance with the recommendations of the American Psychological Association's Ethical Principles for Psychologists and Code of Conduct. All participants gave informed consent in accordance with the Declaration of Helsinki before participating in the study, with the understanding that they could withdraw at any time. There were no ethical concerns about the proposed experiments that required further review.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agresti, A. (2013). Categorical data analysis (3rd ed.). Wiley. https://doi.org/10.1111/biom.12128

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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*(9), 1417. https://doi.org/10.1037/xlm0000246

Butler, A.C., Roediger, H.L. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition* 36, 604–616 (2008). https://doi.org/10.3758/MC.36.3.604

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology* 1, 496–511. https://doi.org/10.1038/s44159-022-00089-1

Carvalho, P., Chen, C.-H., & Yu, C. (2021). The distributional properties of exemplars affect category learning and generalization. *Scientific Reports*, *11*, Article 11263. https://doi.org/10.1038/s41598-021-90743-0

Carvalho, M., Cooper, A., & Marmurek, H. H. C. (2023). Covert retrieval yields a forward testing effect across levels of successive list similarity. *Metacognition Learning*, *18*, 847–861. https://doi.org/10.1007/s11409-023-09348-9

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699. https://doi.org/10.1037/xlm0000406

Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language*, *102*, 83–96. https://www.doi.org/10.1016/j.jml.2018.05.007

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*(11), 1111–1146. https://doi.org/10.1037/bul0000166

Chang, M., & Brainerd, C. J. (2024). Judgments of learning reactivity on item-specific and relational processing. *Journal of Intelligence*, *12*(1), 1–14. https://doi.org/10.3390/jintelligence12010004

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology*, *70*, 1211–1235. https://doi.org/10.1080/17470218.2016.1175485

Cho, K. W., & Powers, A. (2019). Testing enhances both memorization and conceptual learning of categorical materials. *Journal of Applied Research in Memory and Cognition*, *8*(2), 166–177. https://doi.org/10.1016/j.jarmac.2019.01.003

Choi, S., & Lee, S. W. (2020). Enhancing teacher self-efficacy in multicultural classrooms and school climate: The role of professional development in multicultural education in the United States and South Korea. *AERA Open*, *6*(4), 1–16.

https://doi.org/10.1177/2332858420973574

Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology, 10*, 2755. https://doi.org/10.3389/fpsyg.2019.02755

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory, 26*(1), 105–118. https://doi.org/10.1080/09658211.2017.1404111

Double, K. S., Tran, D., & Goldwater, M. B. (2025). Judgments of learning impair rule-based discovery. *Memory & Cognition*, 1–14. https://doi.org/10.3758/s13421-025-01737-6

Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? A comment on and replication attempt of Dougherty, Scheck, Nelson, and Narens (2005). *Memory & Cognition, 46*(4), 558–565. https://doi.org/10.3758/s13421-018-0791-y

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition, 33*(6), 1096–1115. https://doi.org/10.3758/BF03193216

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374–380. https://doi.org/10.3758/BF03210921

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2013) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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(2), 175–191. https://doi.org/10.3758/BF03193146

Fazendeiro, T., Winkielman, P., Luo, C., & Lorah, C. (2005). False recognition across meaning, language, and stimulus format: Conceptual relatedness and the feeling of familiarity. *Memory & Cognition 33*(2), 249–260. https://doi.org/10.3758/BF03195314

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 632. https://doi.org/10.1037/a0026215

Gouravajhala, R., Wahlheim, C. N., & McDaniel, M. A. (2020). Individual and age differences in block-by-block dynamics of category learning strategies. *Quarterly Journal of Experimental Psychology*, *73*(4), 578–593. https://doi.org/10.1177/1747021819892584

Guzdial, M., & Ericson, B. (2016). *Introduction to computing and programming in Python: A multimedia approach* (3rd ed.). Pearson.

Ha, H., & Lee, H.S. (2024). Evaluating the judgment of learning: Its limited impact and the power of retrieval on inductive learning. *Educational Psychology Review 36*, 60. https://doi.org/10.1007/s10648-024-09899-6

Hausman, H., & Kubik, V. (2023). Delayed metacomprehension judgments do not directly improve learning from texts. *Journal of Intelligence, 11*(7), 150. https://doi.org/10.3390/jintelligence11070150

Hausman, H., Kubik, V., & Rhodes, M. G. (2025). *Does test difficulty moderate the forward testing effect?* [Preregistration]. Open Science Framework. https://osf.io/jfxn6/overview?view_only=d93bf4c76a05487fa540756e634aab9b

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1989). *Induction: Processes of* *inference, learning, and discovery*. MIT Press.

Hui, L., de Bruin, A. B., Donkers, J., & van Merriënboer, J. J. (2022). Why students do (or do not) choose retrieval practice: Their perceptions of mental effort during task performance matter. *Applied Cognitive Psychology*, *36*(2), 433–444. https://doi.org/10.1002/acp.3933

Ingendahl, F., Halamish, V., & Undorf, M. (2025). Do immediate judgments of learning alter memory performance? A meta-analytical review. *Psychological Bulletin, 151*(7), 892–929. https://doi.org/10.1037/bul0000487

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1441. https://doi.org/10.1037/a0020636

Kang, Y., Ha, H., & Lee, H. S. (2023). When more is not better: Effects of interim testing and feature highlighting in natural category learning. *Educational Psychology Review*, *35*(2), Article 51. https://doi.org/10.1007/s10648-023-09772-y

Kaplan, A., & Murphy, G. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 829–846. https://doi.org/10.1037/0278-7393.26.4.829

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31*(6), 918–929. https://doi.org/10.3758/BF03196445

Klar, M., Buchner, J. & Kerres, M. (2024). Limits of Metacognitive Prompts for Confidence Judgments in an Interactive Learning Environment. *Open Education Studies*, *6*(1), 20220209. https://doi.org/10.1515/edu-2022-0209

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

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

Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology, 110*(2), 203–217. https://doi.org/10.1037/edu0000211

Lee, H. S., & Ha, H. (2019). Metacognitive judgments of prior material facilitate the learning of new material: The forward effect of metacognitive judgments in inductive learning. *Journal of Educational Psychology, 111*(7), 1189–1201. https://doi.org/10.1037/edu0000339

Lenth, R. V. (2025). *emmeans: Estimated marginal means, aka least-squares means*. [R package] (Version 1.11.2-80003). https://github.com/rvlenth/emmeans

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, 31*(7), 918–930. https://doi.org/10.1080/09658211.2023.2208792

Maxwell, N. P. (2025). Judgments of learning improve memory for word lists via enhanced item-specific encoding: Evidence from categorised, uncategorised, and DRM lists. *Memory, 33*(4), 375–389. https://doi.org/10.1080/09658211.2025.2458813

McDaniel, M. A., Cahill, M. J., Frey, R. F., Limeri, L. B., & Lemons, P. P. (2022). Learning introductory biology: Students' concept-building approaches predict transfer on biology exams. *CBE—Life Sciences Education*, *21*(4), ar65. https://doi.org/10.1187/cbe.21-12-0335

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1084–1097. https://doi.org/10.1037/a0012580

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition*, *29*, 131–140. https://doi.org/10.1016/j.concog.2014.08.008

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

Miyatsu, T., & McDaniel, M. A. (2019). Adding the keyword mnemonic to retrieval practice: A potent combination for foreign language vocabulary learning? *Memory & Cognition*, *47*(7), 1328–1343. https://doi.org/10.3758/s13421-019-00936-2

Murphy, G. L. (2004). *The big book of concepts*. MIT Press. https://doi.org/10.7551/mitpress/1602.001.0001

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*(2), 530–556. https://doi.org/10.3758/s13428-017-0884-8

Ozuru, Y., Kurby, C. A., & McNamara, D. S. (2012). The effect of metacomprehension judgment task on comprehension monitoring and metacognitive accuracy. *Metacognition Learning 7*(2), 113–131. https://doi.org/10.1007/s11409-012-9087-y

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K. H. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 287–297. https://doi.org/10.1037/a0021801

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association, 88*(424), 1273–1283. https://doi.org/10.1080/01621459.1993.10476408

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

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 & Review*, *30*(2), 676–687. https://doi.org/10.3758/s13423-022-02174-1

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

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition*, *33*(6), 1116–1129. https://doi.org/10.3758/BF03193217

Stürmer, S., Christ, O., Jonkmann, K., Josephs, I., Gaschler, R., Glöckner, A., Mokros, A., Rohmann, A., & Salewski, C. (2018). 10 Jahre universitäres Fernstudium in Psychologie an der FernUniversität in Hagen. *Psychologische Rundschau*, *69*, 104–108, https://doi.org/10.1026/0033-3042/a000400

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*(4), 331–362. https://doi.org/10.1080/01638530902959927

Vaughn, K. E., Hausman, H., & Kornell, N. (2017). Retrieval attempts enhance learning regardless of time spent trying to retrieve. *Memory*, *25*(3), 298–316. https://doi.org/10.1080/09658211.2016.1170152

Wang, X., Liu, X., Chen, L., Feng, K., Ye, Q., & Zhu, H. (2023). The forward effect of delayed judgments of learning is influenced by difficulty in memory and category learning. *Journal of Intelligence*, *11*(6), 101. https://doi.org/10.3390/jintelligence11060101

Wiens, S., & Nilsson, M. E. (2017). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. *Educational and Psychological Measurement*, *77*(4), 690–715. https://doi.org/10.1177/0013164416668950

Yang, C., Luo, L., Vadillo, M. A., Yu, R. & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 339–435. https://doi.org/10.1037/bul0000309

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *Npj Science of Learning, 3*(1), 8. https://doi.org/10.1038/s41539-018-0024-y

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 485. https://doi.org/10.1037/xlm0000449

Yang, C., Zhao, W., Luo, L., Sun, B., Potts, R., & Shanks, D. R. (2022). Testing potential mechanisms underlying test-potentiated new learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(8), 1127–1143. https://doi.org/10.1037/xlm0001021

Yu, Y., Zhao, W., Li, A., Zhang, Y., & Yang, C. (2025). Is covert retrieval an effective learning strategy? Is it as effective as overt retrieval? Answers from a meta-analytic review. *Educational Psychology Review, 37*(52). https://doi.org/10.1007/s10648-025-10024-4

Zhang, X., Guo, Y., Jiang, Y., & Yuan, Y. (2024). Judgments of learning reactively affect memory by inducing covert retrieval. *The Journal of General Psychology*, 1–29. https://doi.org/10.1080/00221309.2024.2409785

Zhao, W., Xu, M., Xu, C., Li, B., Hu, X., Yang, C., & Luo, L. (2023). Judgments of learning following retrieval practice produce minimal reactivity effect on learning of education-related materials. *Journal of Intelligence, 11*(10), 190. https://doi.org/10.3390/jintelligence11100190

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An Overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2