How and When Judgements of Learning Facilitate Inductive Learning

Morgan D. Shumaker

Radford University

A thesis submitted to the faculty of Radford University
in partial fulfillment of the requirements for the degree
of Master of Arts in the Department of Psychology

Thesis Advisor: Kathleen M. Arnold, Ph.D.

May 2021

Kathleen M. Arnold
Dr. Kathleen M. Arnold
Thesis Advisor

Catherine Middlebrooks-Evans
Dr. Catherine Middlebrooks-Evans
Committee Member

Thomas Pierce
Dr. Thomas Pierce
Committee Member

Abstract

Students often have to learn inductively by studying individual examples and learning to abstract an overarching rule. Judgments of learning (JOLs)—metacognitive self-judgments of the likelihood of remembering information in the future—may also facilitate subsequent inductive learning. Specifically, making JOLs on previously learned categories may facilitate subsequent inductive learning of new categories. However, this finding has only occurred under circumstances in which the JOLs provided a retrieval opportunity and focused on the category as a whole, rather than on individual examples of a category. This study aimed to parse the roles of retrieval and type of encoding processes (relational vs. item-level) in the effect of JOLs on facilitating subsequent inductive learning by manipulating the type of JOL and category salience. Participants studied paintings from various artists and made either cue-only or cue-target JOLs under conditions in which the category was either salient or not salient, prior to studying a new set of artists and taking a final test. The results indicate that making cue-only JOLs may facilitate inductive learning for new categories, but only under circumstances in which the overarching category is salient.

Morgan D. Shumaker, M.A.
Department of Psychology, 2021
Radford University

TABLE OF CONTENTS

List of Figures

Chapter 1: Introduction

When learning, people are often required to learn conceptual information. That is, rather than memorizing specific instances, individuals must learn broader concepts that are not explicitly stated and can be applied to new situations. This type of learning, known as inductive learning, is achieved by abstracting rules from examples of the to-be-learned concept (Kornell & Bjork, 2008). Inductive learning is particularly useful in category learning. For example, a course may require students to differentiate between plant species. By studying exemplars (i.e., category members), students can learn to abstract rules to later determine to which species a new plant belongs (Brunmair & Richter, 2019). Prior research suggests that two potential ways to enhance inductive learning are through the use of retrieval practice and metacognitive judgments of learning (Jacoby et al., 2010; Lee & Ha, 2019).

Effects of Retrieval

The benefits of retrieval are especially pertinent to academia, as testing is a common means of assessment and involves an active learning process that enhances long-term retention (Kornell et al., 2009; Rowland, 2014). Several studies have demonstrated the benefits of retrieval practice in many different contexts. For example, retrieval enhances the memory and retention of material, a finding known as the testing effect (Roediger & Karpicke, 2006). One theory for the testing effect is that retrieval may strengthen subsequent storage and slow forgetting of the tested material (Kornell & Bjork, 2007; Roediger & Karpicke, 2006; Storm & Bjork, 2010). Likewise, initially retrieving information from long-term memory leads to that information becoming more accessible in the future (Storm & Bjork, 2010). Retrieval can also involve deep processing, which subsequently leads to deeper learning (Craik & Tulving, 1975; McDaniel et al., 2009). Another theory is that retrieving information related to the target may increase the number of retrieval routes or strengthen existing retrieval routes, thereby leading to an increased probability of retrieving the target in the future (Butler, 2010; Butler et al., 2017; McDaniel et al., 1989). In this vein, by retrieving other studied information, that information may become integrated and give way to an updated memorial representation of the learned material (Butler et al., 2017; Carpenter, 2009).

Retrieval practice can also enhance inductive learning (Jacoby et al., 2010). In one experiment, testing students on previously studied exemplars of individual bird families enhanced those students' abilities to classify previously studied exemplars, a traditional testing effect. Moreover, testing on previously studied exemplars enhanced students' abilities to classify novel (i.e., previously unstudied) exemplars of the previously studied categories, showing that testing also enhances inductive learning. Jacoby et al. (2010) suggest that retrieval facilitates inductive learning through a series of steps. First, retrieval enhances the learning of previously studied exemplars by enhancing memory for the distinctive features between members of different categories. Second, retrieval encourages students to discover features that are shared by members of a category, or are not shared between members of different categories, allowing for rule abstraction (see also Kornell & Bjork, 2008).

In addition to enhancing memory for previously studied information, taking a test on one set of material can enhance learning for a subsequently studied set of new material, even when the second set of material is unrelated to the first. This finding, termed test-potentiated-new-learning (TPNL), has been demonstrated in a variety of materials ranging from paired associates to short passages (Chan, Meissner, & Davis, 2018). One theory of TPNL suggests that taking a test between sets of material increases the cognitive resources one has available to encode new information by protecting learners from proactive interference, in which the build-up of prior information negatively impacts one's ability to remember subsequently encoded information (Bennett, 1975; Chan, Manley, et al., 2018). Szpunar et al. (2013) proposed another hypothesis that retrieval practice reduces mind wandering and allows learners to refocus their attention before encoding the new set of material. Other research suggests that TPNL is driven by the metacognitive benefits of retrieval: specifically, by informing students about what they do or do not know. Additionally, retrieval practice may also encourage learners to hypothesize about the type of information they will be tested on later or about future testing formats, allowing them to adjust their subsequent encoding strategies accordingly (Chan, Manley, et al., 2018; Chan, Meissner, & Davis, 2018). Thus, testing may potentiate learners' ability to learn new information by enhancing their metacognitive awareness and/or through non-metacognitive consequences of retrieval practice, or both.

Experience with prior testing may also facilitate inductive learning for new concepts. For example, in an experiment by Lee and Ahn (2018), participants were asked to study paintings by various artists and then take a test, restudy the paintings, or complete a distractor task before moving on to a second set of paintings by new artists. The researchers were particularly interested in how taking the test on the first set of artists impacted participants' abilities to learn the second set of artists. They found that participants who took a test after studying the first set of artists were better able to identify which of the second set of artists painted both previously studied paintings and new paintings than the other two groups. In other words, testing between sets of artists enhanced participants' inductive learning of the second artists' painting styles. Lee and Ahn (2018) speculated that after making this assessment, students may have subsequently modified their study strategies in a way that was better suited for learning a concept (e.g., finding similarities between stimuli) and generalizing that concept to other stimuli.

Item-specific vs. Relational Processing

Another factor that may impact inductive learning is the type of processing engaged by the learners, which can be influenced by category salience. Prior research suggests that category salience plays a role in the ability to learn inductively by influencing the way information is processed. For example, when learning inductively in situations where the categorical relationship between exemplars is salient, learners tend to abstract an overarching rule that allows them to correctly classify new exemplars of the studied categories (Goldwater et al., 2018). Moreover, in circumstances where the categorical relationship is salient, learners tend to engage in relational processing by encoding the shared features between items of the same category (McDaniel et al., 2015).

Conversely, when categorical relationships are not salient, learners tend to memorize particular examples rather than abstracting a rule, resulting in a lesser ability to correctly classify new exemplars than those who abstract a rule (Goldwater et al., 2018). In this vein, the lack of category salience can cause learners to engage in item-specific processing by encoding features that are specific to each example (McDaniel et al., 2015). Thus, by manipulating the saliency of a category, researchers can manipulate the type of processing that learners engage in when encoding information, leading to differences in the information recalled (McDaniel et al., 2015). Further, learners can learn to switch from item-specific processing to relational processing and vice versa, based on the goal of the task (Goldwater et al., 2018).

Metacognition and Judgments of Learning

Despite the multitude of learning benefits that retrieval practice may provide, it is not always a tool that students use for learning. That is, whereas students take tests to demonstrate or assess their knowledge, they do not necessarily test themselves in order *increase* their knowledge

(Karpicke et al., 2009). Because students are frequently required to self-regulate their learning outside of the classroom, they must make decisions about what to study and how long to study (Middlebrooks & Castel, 2018). Learners engage in metacognitive monitoring—assessing their current state of knowledge—and metacognitive control—the regulation of cognitive activity—when regulating their studying (Dunlosky & Metcalfe, 2009; Dunlosky & Tauber, 2016; Metcalfe & Finn, 2008a; Rhodes & Castel, 2008; Rhodes & Tauber, 2011).

Finding a way to enhance students' willingness to use retrieval-practice as a means of studying may benefit their learning and allow them to better succeed. One way to do this is by incorporating retrieval-practice into strategies students already use—metacognitive assessments. Consider a situation in which students are studying for an upcoming exam. The students might first determine how well they have learned each section of material. Following this method of metacognitive monitoring, those students may use their metacognitive control to allocate more study time toward the material they think they have learned the least (Kornell & Bjork, 2007; Metcalfe, 2009). Making judgments of learning (JOLs), judgments about the likelihood of remembering studied items on a future test, is one specific tool that students may use to better regulate their learning.

For JOLs to be an effective study tool, they must be accurate. If inaccurate JOLs are used to guide studying behavior, the studying behavior will be ineffective (Metcalfe & Finn, 2008a; Nelson & Dunlosky, 1991). Fortunately, when making JOLs, learners can sometimes correctly predict their future performance (Koriat, 2019). Under certain circumstances, such as when learning categories, learners can correctly predict the probability of recalling certain items relative to other items, a type of accuracy known as relative accuracy or resolution (Dunlosky & Metcalfe, 2009; Dunlosky & Nelson, 1992; Jacoby et al., 2010; Koriat, 2008; Little & McDaniel, 2015b). In other words, they can sometimes predict which items are more likely or less likely to be recalled. Additionally, learners can sometimes correctly predict the overall proportion of items they will be able to recall, another type of accuracy termed absolute accuracy or calibration (Koriat, 2019; Scheck et al., 2004).

Accurate JOLs seem to be limited to situations in which the judgment is delayed and only the cue is available (meaning the target itself must be retrieved). When compared to making JOLs immediately after studying an item, learners' judgments are more accurate when the JOL is made after all the items have been studied, a finding termed the delayed-JOL effect (Dunlosky & Metcalfe, 2009; Dunlosky & Tauber, 2016; Koriat & Bjork, 2005; Metcalfe & Finn, 2008b; Rhodes & Tauber, 2011; Scheck et al., 2004). Additionally, when compared to making JOLs in a situation where both the cue and the target are present, JOLs in situations where only the cue is present are more accurate. This is likely because the delayed cue-only JOL serves as an indication of learners' abilities to retrieve the target from long-term memory, making it more diagnostic of their ability to retrieve it on a future test.

Making delayed cue-only JOLs may have benefits beyond better accuracy. For example, the memory hypothesis suggests that retrieving the target directly affects memory by increasing the target's accessibility. Consequently, the target becomes more retrievable in the future because retrieving the target during the JOL serves as a source of spaced retrieval-practice (Kimball & Metcalfe, 2003; Spellman & Bjork, 1992). Further, delayed cue-only JOLs may allow students to diagnose a gap in their knowledge by comparing their current knowledge state to their desired knowledge state (Metcalfe, 2009). Learners can, in general, typically determine whether or not something has been stored in their memory after making an unsuccessful retrieval attempt (Cantor et al., 2015). In other words, learners tend to know whether the required knowledge is available, although currently inaccessible, or whether it is both unavailable and inaccessible (Cantor et al., 2015).

There has been some research examining whether the act of making JOLs can directly enhance learning. Soderstrom et al. (2015) found that when learning unrelated, weakly-related, and strongly-related word pairs, participants who made cue-target JOLs after studying, had better memory for strongly-related (but not weakly-related or unrelated) word pairs than those who just restudied. Other studies have also shown that making a JOL can enhance memory for material similar to the testing effect, but only when the JOL involves a retrieval opportunity (cue-only) (see Dougherty et al., 2005; Jönsson et al., 2012). In these instances, the act of making a JOL may result in altering the memory representation of the judged material, making it easier to recall at a later point in time (Jönsson et al., 2012; Tekin & Roediger, 2021).

Facilitating Inductive Learning

Because retrieval can facilitate inductive learning and certain types of JOLs can enhance learning, Lee and Ha (2019) investigated whether making JOLs could facilitate subsequent inductive learning. Additionally, Lee and Ha investigated how the type of processing encouraged by the JOL may affect inductive learning by testing two types of JOLs. Specifically, Lee and Ha (2019) examined two types of JOLs: item-level judgements (ILJs), which focus learners' attention on one specific example from a category, and category-level judgments (CLJs), which focus learners' attention to the category as a whole, rather than on one example. In one experiment, participants studied paintings by various artists and then took a cued-recall test, restudied the artist-painting pairs, or made JOLs on those paintings. JOLs were made at the item-level, in which participants were shown each painting with the artist's name (cue-target) again and asked to predict the likelihood of being able to correctly classify new paintings by that artist.

After this, they studied a new set of artist-painting pairs, followed by taking a final test. On the final test, which consisted of new paintings by artists from both studied sets, participants across the three conditions (ILJ, test, or restudy) performed equally well on classifying paintings from the first set; however, participants who took a test after the first set outperformed the ILJ and restudy groups when classifying novel paintings from the second set of artists. Moreover, there was no difference in performance between those who made ILJs and those who restudied, suggesting that ILJs did not facilitate inductive learning for the second set of artists.

In a second experiment, Lee and Ha (2019) had participants go through the same procedure, the only change being that participants made CLJs rather than ILJs. When making the CLJs, participants were asked the same question as in Experiment 1, but were only shown the name of the artist when making their judgments. Mirroring the results from Experiment 1, there was no difference between the groups when classifying novel paintings from the first set of artists. When classifying novel paintings from the second set, however, both the testing group and the CLJ group outperformed the restudy group. Further, there was no difference in performance between the testing group and CLJ group, suggesting that making CLJs may be as effective as testing in facilitating inductive learning (Lee & Ha, 2019).

Lee and Ha (2019) suggest that retrieval may be the factor driving the difference between ILJs and CLJs in facilitating inductive learning because the CLJs provided a retrieval opportunity, whereas the ILJs did not. The retrieval opportunity when making the CLJs—but not the ILJs—may have allowed participants to receive a benefit similar to test-potentiated-new-learning. A second possibility is that making CLJs focused participants’ attention on the broader category (i.e., painting styles), while making the ILJs focused participants’ attention on the specific items (i.e., the individual paintings). Focusing learners’ attention on the broader concept may have encouraged them to adjust their study strategy for the second set of artists and to abstract a category rule to classify the new paintings. In contrast, making JOLs on the first set of paintings as individual items rather than as exemplars of the larger category may not encourage such rule abstraction, thereby limiting inductive learning and subsequent transfer (Lee & Ha, 2019). Thus, the differences in performance may be a result of differences in the way the second set of artists was processed, owing to the way in which the first set of artists was judged.

Chapter 2: Current Study

The purpose of the current study is to separate the effects of retrieval and type of processing on inductive learning. Specifically, where Lee and Ha (2019) confounded retrieval and type of processing, I separated these variables in a crossed design. I manipulated the type of JOL (i.e., cue-only vs. cue-target) in order to determine whether or not including a retrieval opportunity is necessary to facilitate inductive learning. Additionally, I investigated the role of encoding processes on facilitating inductive learning by manipulating category salience (i.e., salient vs. not salient); specifically, I used category salience while making the JOL to encourage either relational or item-level processing during subsequent study. When the category is salient, participants might engage in relational processing and when the category is not salient, participants might engage in item-level processing (McDaniel et al., 2015). Finally, I measured both relative and absolute JOL accuracy to gain insight into learners' abilities to correctly assess and predict their inductive learning.

In the current study, participants studied one set of artists (Section A) and then made either cue-only or cue-target JOLs on either three paintings (category salient) or one painting (category not salient) from each artist during the JOL. After making these JOLs, participants studied a second set of artists (Section B) before taking a final test on new paintings from all of the artists in both sections. The primary focus of the present study was inductive learning on Section B because the circumstances in which participants' make their JOLs should influence how they approach studying the second set of artists.

Because retrieval can facilitate inductive learning (Lee & Ahn, 2018), I expected to find a main effect of type of JOL on inductive learning for Section B (post-JOL) material (Butler, 2017; Jacoby et al., 2010). Specifically, those that make cue-only JOLs on Section A material should display better inductive learning on Section B material than those who make cue-target JOLs. I also expected an interaction between JOL and Category Salience in that cue-only JOLs on Section A material will only impact inductive learning of Section B material when the category is salient. Therefore, those in the salient cue-only group will outperform those in the salient cue-target group on Section B material. Making a category-salient JOL may shift learners' attention to the similarities and differences between categories (i.e., artists' painting styles), encouraging relational encoding when studying new material (McDaniel et al., 2015). In contrast, making a JOL when the categories are not salient may shift learners' attention to item-specific information (i.e., individual paintings) and encourage item-focused processing of new material (McDaniel et al., 2015). However, I expected no difference between the cue-only and cue-target conditions when the category was not salient. That is, inductive learning should be equivalent between those in the not salient cue-only group and those in the not salient cue-target group; without the retrieval opportunity, learners may not benefit from making either type of judgment (Lee & Ha, 2019; Rhodes & Tauber, 2011), regardless of their encoding processes (item-specific vs. relational). Finally, there should be no difference in inductive learning between any of the four groups on Section A (pre-JOL) material, as cue-only JOLs and category salient conditions should enhance subsequently encoded information (Section B) rather than previously encoded information (Lee & Ahn, 2018; Lee & Ha, 2019).

Chapter 3: Methods

Participants

Participants consisted of 204 undergraduate psychology students from Radford University. Participants ranged from 18-38 years old (*M* = 19.83, *SD* = 2.87) and consisted of 46 males, 149 females, two who identified as other, and seven who preferred to not specify. They were recruited online via Radford University's SONA system and by asking psychology department instructors to offer course extra credit to their students for participating. All participants received partial course credit as compensation. According to a post-hoc power analysis using G*Power (Faul et al., 2007) to detect a medium effect (η_p^2 = .09) for a between-subject analysis of variance with four groups, this study reached 94% power.

Design

This experiment utilized a 2 (Category Salience: Salient vs. Not salient) x 2 (JOL: Cue-only vs. Cue-target) between-subjects design. The first dependent variable is the proportion of new paintings correctly matched to the artist that created it, conceptualized as a measure of inductive learning. The second dependent variable is relative accuracy, measured by the correlation between participants' JOL for each artist and the average test performance for the corresponding artist. Finally, absolute accuracy, often called calibration, was measured by calculating the difference scores between each participant's average JOL and average overall test performance.

Materials

Stimuli included 108 paintings of landscapes created by 12 different artists, adopted from Kornell and Bjork (2008). Paintings were split in half between two study sessions (Set 1 and Set 2) with each section containing six paintings by six artists for a total of 36 paintings. In both study phases, paintings were presented using blocked randomization with one block containing one painting by each of the six artists for a total of six blocks such that the blocks were unique for each participant. Specifically, one painting from each of the six artists in Set 1 were presented in the first block and a new painting by each of the six artists in Set 2 were presented in each of the subsequent five blocks. Set 2 was presented in the same fashion and contained paintings from six new artists (i.e., artists not featured in Phase 1). Set 1 and Set 2 were counterbalanced between study phases such that half of the participants studied Set 1 first, while the other half studied Set 2 first.

The final test was administered in a multiple-choice format and consisted of three new (i.e., previously unstudied) paintings from each of the 12 studied artists. Participants were presented with each of the new paintings and asked to select which of the 12 artists created each painting from a list of all 12 artists' names.

Procedure

Participants completed the study online through Qualtrics and were randomly assigned to one of four conditions: Salient cue-only (S-CO), Salient cue-target (S-CT), Not salient cue-only (NS-CO), and Not salient cue-target (NS-CT). A schematic representation of the procedure is provided in Figure 1. Participants began studying the first set of six artists across six blocks, totaling 36 paintings. Each painting along with the artist's name was presented for three seconds, with a one second blank screen between each artist-painting pair.

After studying the first set of artist-painting pairs, participants were asked to make a series of six JOLs (one per artist). While making their JOLs, those in the S-CO condition saw three previously studied paintings (at once) from each artist, without the artist's name present (i.e., cue-only). The S-CT condition saw three previously studied paintings from each artist with the artist's name displayed under the paintings (i.e., cue-target). The NS-CO condition saw one painting from each of the six artists, without the artist's name present (i.e., cue-only). Finally, the NS-CT condition was shown one painting from each of the six artists, with the artist's name present (i.e., cue-target). Each of the six JOLs were made individually (see Figure 2). Participants were asked to respond to "What is the likelihood you would later be able to correctly identify a new painting by this artist?" by indicating the likelihood on a sliding scale ranging from 0 (*not at all likely*) to 100 (*very likely*), located directly under the painting(s). All participants were instructed to interpret the scale as a percent value.

Following the six JOLs, participants studied the second set of six artists in the same fashion as the previous set. Participants did not make JOLs for the second set of artists. Directly following the study of the second set, participants were given a multiple-choice test that consisted of three new paintings by each of the 12 studied artists (i.e., from Section A and Section B). Each question consisted of a new painting from one of the 12 artists and participants were asked to select the most likely artist of the painting from a list of all 12 names. The list of 12 artists appeared in a fixed order (see Figure 3). Following the content questions, participants responded to a series of questions regarding their study strategies and demographic information.

Figure 1

Schematic representation of the procedure

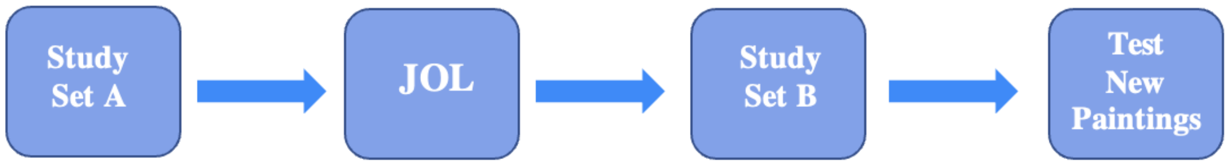
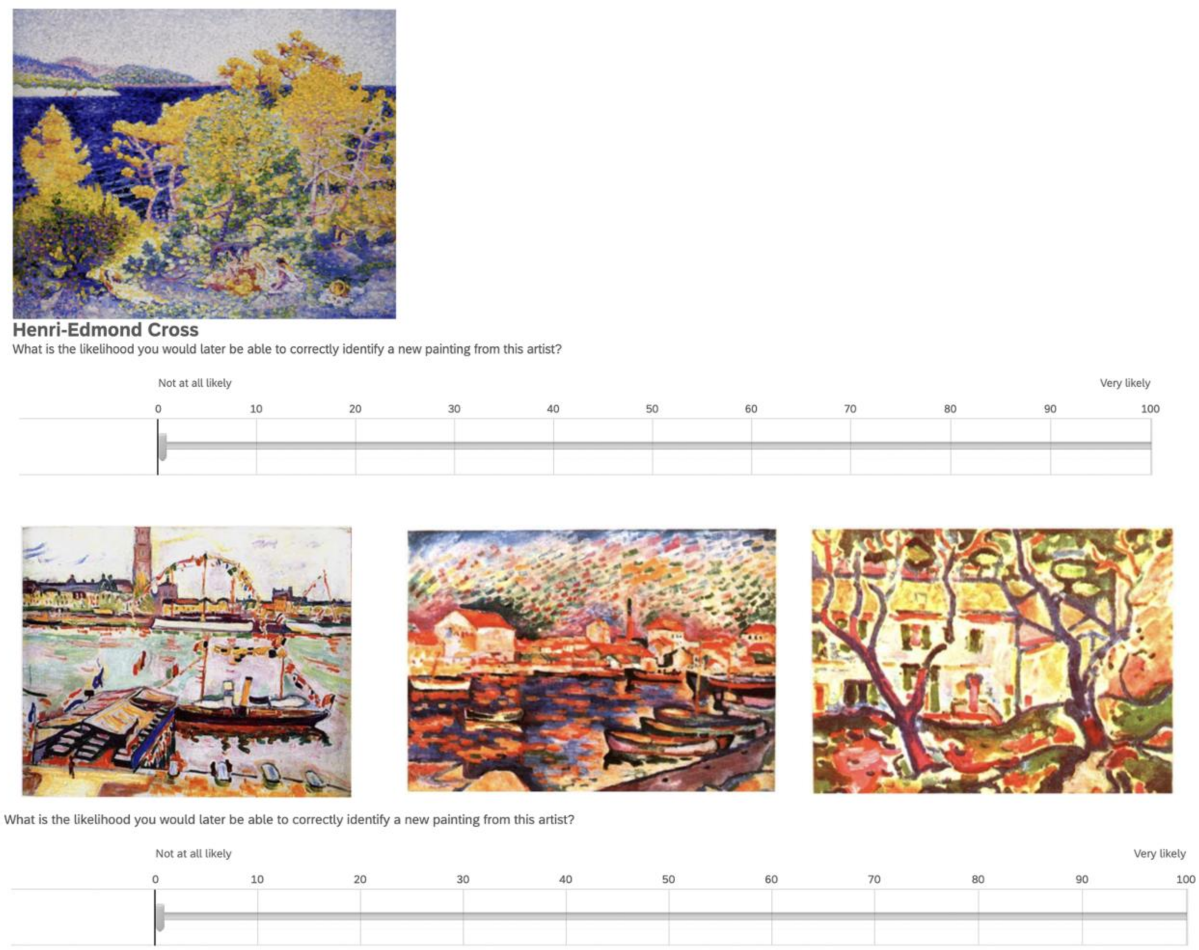


Figure 2

Examples of JOLs



Non-salient cue-target JOL (top) and a Salient cue-only JOL (bottom).

Figure 3

Example of final test question

What is the name of the artist who created this painitng?

- ○ Bruno Pessani
- ○ Cirprian Stratulat
- ○ George Wexler
- ○ Georges Braque
- ○ Georges Seurat
- ○ Henri-Edmond Cross
- ○ Judy Hawkins
- ○ Marilyn Mylrea
- ○ Philip Juras
- ○ Ron Schlorff
- ○ Ryan Lewis
- ○ Yie Mei

Chapter 4: Results

Inductive Learning

Inductive learning is operationalized as the proportion of new paintings from each of the studied artists that participants correctly identified. First, Section B is reported, as the main focus of this study was to determine how making JOLs under different circumstances facilitates inductive learning. Specifically, the effects of JOL type and Category Salience on learning in Section B are examined. Then, performance on Section A is examined to confirm that making JOLs enhances subsequently encoded information, rather than previously encoded information.

Section B

A 2 (JOL: cue-only vs. cue-target) x 2 (Category Salience: salient vs. not salient) between-subjects ANOVA was conducted to analyze differences in performance on Section B (post-JOL) material. As predicted, those in the cue-only condition (M = .26) correctly identified more of the new (unstudied) paintings in Section B than those in the cue-target condition (M = .21) (see Figure 5). However, this main effect was only marginally significant, $F(1, 202) = 3.19$, $\eta_p^2 = .016$, $p = .076$. There was no main effect of Category Salience on inductive learning for Section B material, suggesting that category salience, alone, does not influence inductive learning, $F < 1$.

However, there was a marginal interaction between JOL and Category Salience in the predicted direction, $F(1, 202) = 3.00$, $\eta_p^2 = .015$, $p = .085$. This finding suggests that the marginal effect of JOL may be dependent on category salience. An independent *t*-test revealed that when the category was salient, those in the cue-only group (M = .28, SD = .23) outperformed those in the cue-target group (M = .18, SD = .17), $t(98) = 2.34$, d = .05, p = .02. In contrast, when the category was not salient, there was no difference between the cue-only (M = .23, SD = .19) and cue-target (*M* = .23, *SD* = .22) JOL conditions in correct identification of new paintings by Section B artists, $t < 1$ (see Figure 4). Overall, performance across groups was near floor (eight out of 36 paintings), which may have made it difficult to find significant effects (see Figure 5).

Section A

To analyze performance on Section A (pre-JOL) test material, a 2 (JOL: cue-only vs. cue-target) x 2 (Category Salience: salient vs. not salient) between-subjects ANOVA was conducted to examine the effects of JOL and Category Salience on inductive learning of previously studied material. Consistent with the hypothesis regarding Section A performance, there was no significant effect of JOL type, Category Salience, and no interaction on inductive learning of previously studied material, all *F*s < 1 (see Table 1).

Table 1

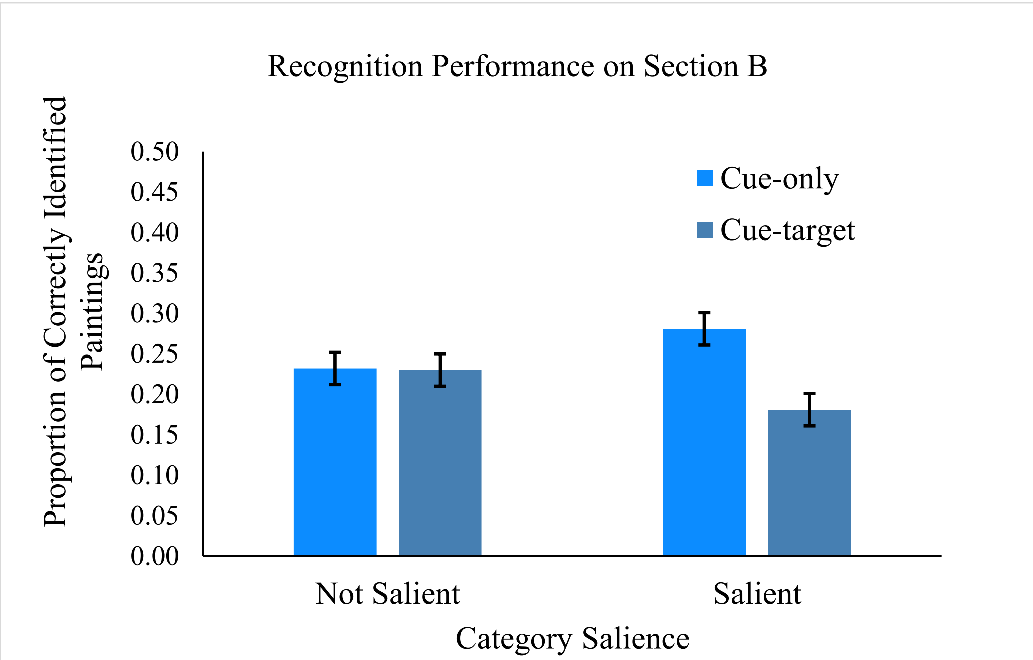
Average performance on Section A and Section B

	Section A		Section B	
	M	*SD*	*M*	*SD*
Cue-only				
S	.23	.19	.28	.23
NS	.20	.15	.23	.19
Cue-target				
S	.17	.16	.18	.17
NS	.20	.20	.23	.22

[a]Section A consists of novel paintings by each of the artists on which participants made JOLs. Section B consists of novel paintings by novel artists, studied after JOLs.

Figure 4

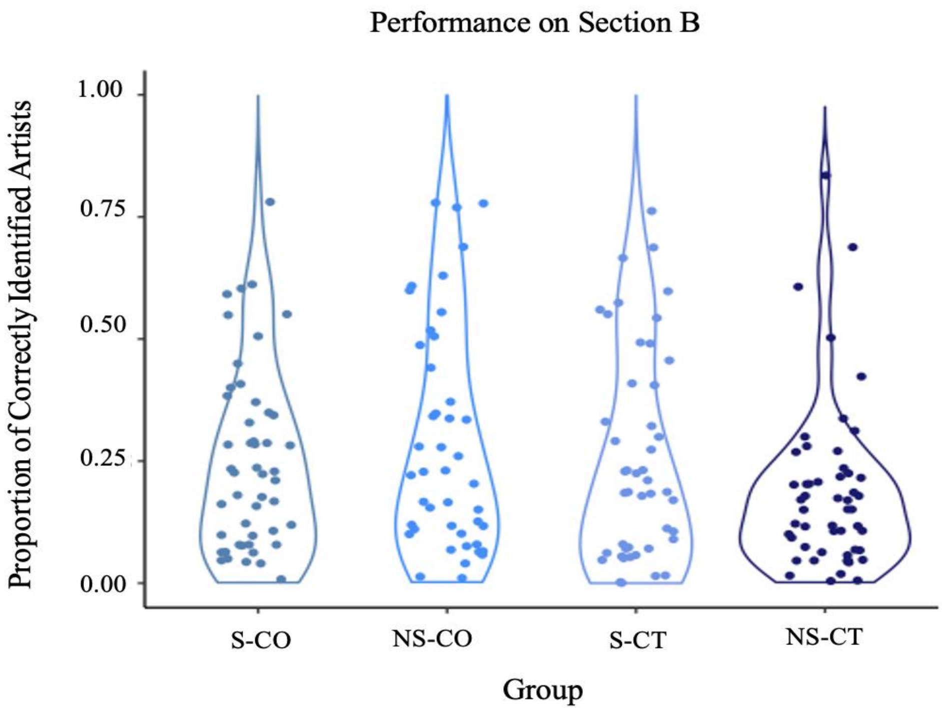
Performance on Section B



The proportion of correctly identified novel paintings by each of the artists that were studied in Section B (post-JOLs). Error bars represent standard error.

Figure 5

Individual scores on Section B



The distribution of performance on Section B, grouped by condition. Markers represent the individual score for each participant.

Metacognitive Accuracy

Metacognitive accuracy was examined in two separate ways. First, resolution was examined by comparing differences in participants' abilities to accurately predict their performance for one artist relative to another artist. Resolution was calculated using both Pearson's product-moment correlations and Gamma correlations to examine the relationship between performance for each artist and participants' JOLs for each artist. Second, calibration was examined by comparing differences in the accuracy of participants' predictions regarding their overall ability to correctly identify new paintings, averaged across artists. Calibration was examined first by calculating each participant's average JOL minus their average performance on Section A. Calibration was also examined by subtracting each participant's performance on each artist from their average JOL for each artist and then, averaging each participant's difference scores to get one difference score for each individual.

Resolution

To examine each participant's ability to accurately assess whether or not they learned one artist's painting style better than another artist's painting style, resolution was computed two ways. First, Pearson product-moment correlations were computed by correlating participants' JOL for each artist with their average performance per artists. A Pearson's product-moment correlation was chosen because JOLs are treated as an interval measure, rather than an ordinal measure, which is consistent with prior literature when examining resolution at the category level (Jacoby et al., 2010). Then, a 2 (JOL: cue-only vs. cue-target) x 2 (Category Salience: salient vs. not salient) between-subjects ANOVA was conducted to examine the effects of JOL and Category Salience on resolution.

Contrary to my hypothesis, there was no main effect of JOL on resolution, $F < 1$. There also was no main effect of Category Salience on resolution, $F(1, 202) = 2.75$, $p > .05$ (see Figure 6). These results suggest that cue-only JOLs do not improve resolution relative to cue-target JOLs, nor does category salience improve resolution relative to when the categorical relationship is not salient. A significant interaction between JOL and Category Salience on resolution was observed, $F(1, 202) = 7.19$, $p = .008$, $\eta_p^2 = .035$. However, the interaction was not in the predicted direction and post hoc analyses using a Bonferroni correction to examine the differences between means failed to reach significance, $p > .05$.

Resolution was also measured using Gamma correlations, by treating JOLs as an ordinal measure, consistent with most prior literature on relative accuracy (Dunlosky & Metcalfe, 2009; Koriat 2009). Specifically, participants' JOLs for each artist were compared with their performance (correct vs. incorrect) on each of the three questions pertaining to each artist. A 2 (JOL: cue-only vs. cue-target) x 2 (Category Salience: salient vs. not salient) between-subjects ANOVA was conducted using Gamma correlations as the dependent variable to examine the effects of JOL and Category Salience on resolution. Consistent with resolution as measured by Pearson's *r*, there was no main effect of JOL on resolution, $F < 1$. There was no main effect of Category Salience on resolution, $F(1, 202) = 1.27$, $p > .05$ (see Table 2). There was a significant interaction between JOL and Category Salience on resolution, $F(1, 202) = 5.76$, $p = .017$, $\eta_p^2 = .028$. However, the interaction was not in the predicted direction and post hoc analyses using a Bonferroni correction to examine the differences between means failed to reach significance, $p > .05$.

Table 2

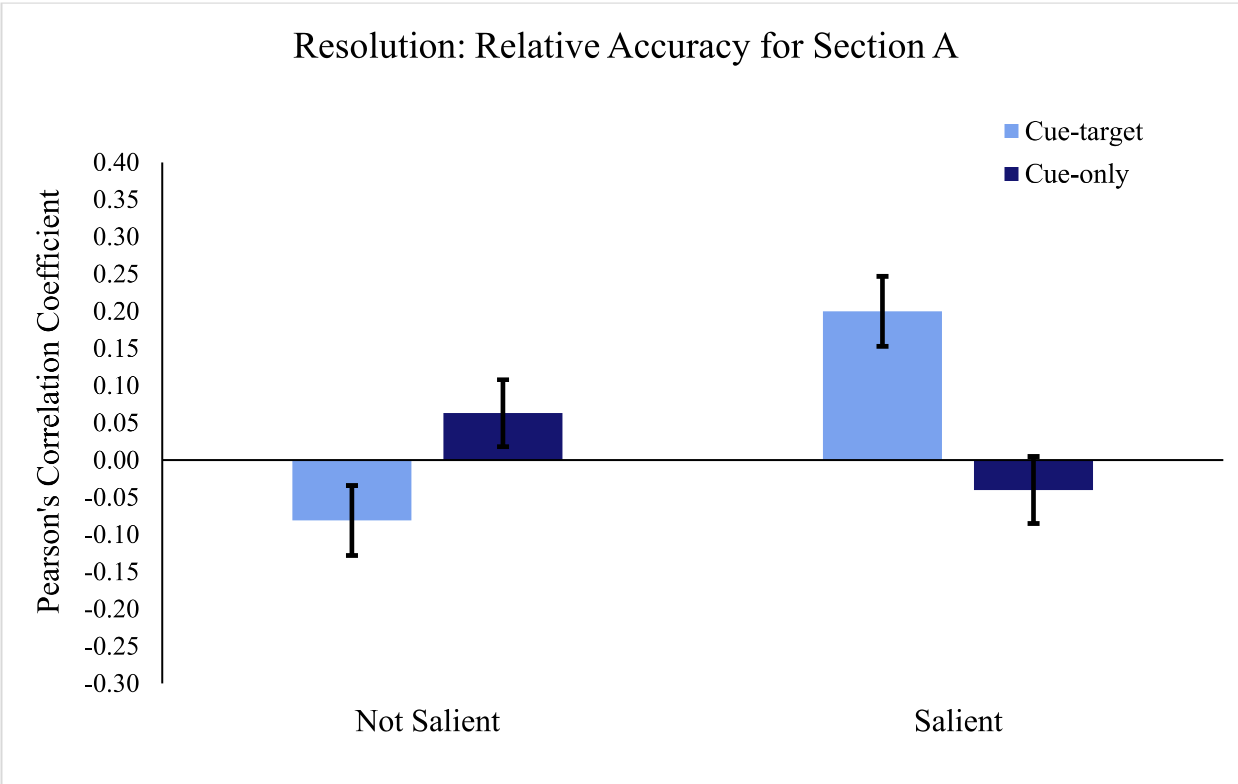
Resolution on Section A

		M	*SEM*
Cue-only			
	S	.03	.08
	NS	.12	.07
Cue-target			
	S	.18	.07
	NS	-.08	.08

[a]Measured by Gamma correlations between participant's performance (correct vs. incorrect) on each artist and their JOL per artist.

Figure 6

Resolution on Section A



Resolution as measured by the correlation between participants' JOL for each individual artist and their average performance for each individual artist (in Section A). Error bars represent standard error.

Calibration

To examine each participant's ability to accurately predict their overall performance on Section A (i.e., across all of the artists for which they made JOLs), difference scores were calculated by subtracting each participant's average JOL by their average performance on Section A (Dunlosky & Metcalfe, 2009). A difference score of zero would indicate perfect calibration (i.e., incredibly accurate), while a difference score of ± 1 would indicate no calibration (i.e., not at all accurate). Additionally, positive scores for calibration are indicative of under-confidence, whereas negative scores for calibration are indicative of overconfidence. After the difference scores were calculated for each participant, a 2 (JOL: cue-only vs. cue-target) x 2 (Category Salience: salient vs. not salient) between-subjects ANOVA was conducted to examine the effects of JOL and Category Salience on calibration.

Participants in all groups were, on average, overconfident in their ability to correctly identify novel paintings from the artists on which they made the JOLs (see Figure 7). In contrast with the prediction that making a cue-only JOL would enhance participants' calibration relative to making a cue-target JOL, there was no effect of JOL type on calibration, $F < 1$. That is, those in the cue-only group did not have better calibration than those in the cue-target group (see Figure 7), suggesting that cue-only JOLs do not improve calibration relative than cue-target JOLs. As hypothesized, there was no main effect of Category Salience on calibration and numerical differences were minimal, suggesting that category salience does not improve calibration relative to when the categorical relationships are not salient. Finally, there was no interaction between JOL and Category Salience on calibration, $F < 1$.

Calibration was also analyzed by calculating the difference between each participant's performance on each of the six artists from Section A and subtracting their JOL for that artist. Difference scores for each individual artist were then averaged together to generate a globalized difference score. A 2 (JOL: cue-only vs. cue-target) x 2 (Category Salience: salient vs. not salient) between-subjects ANOVA was used to analyze the effects of JOL and Category Salience on absolute accuracy. The analysis revealed that there was no main effect of JOL on calibration, no main effect of Category Salience on calibration, and no interaction, all Fs < 1 (see Table 3).

Table 3

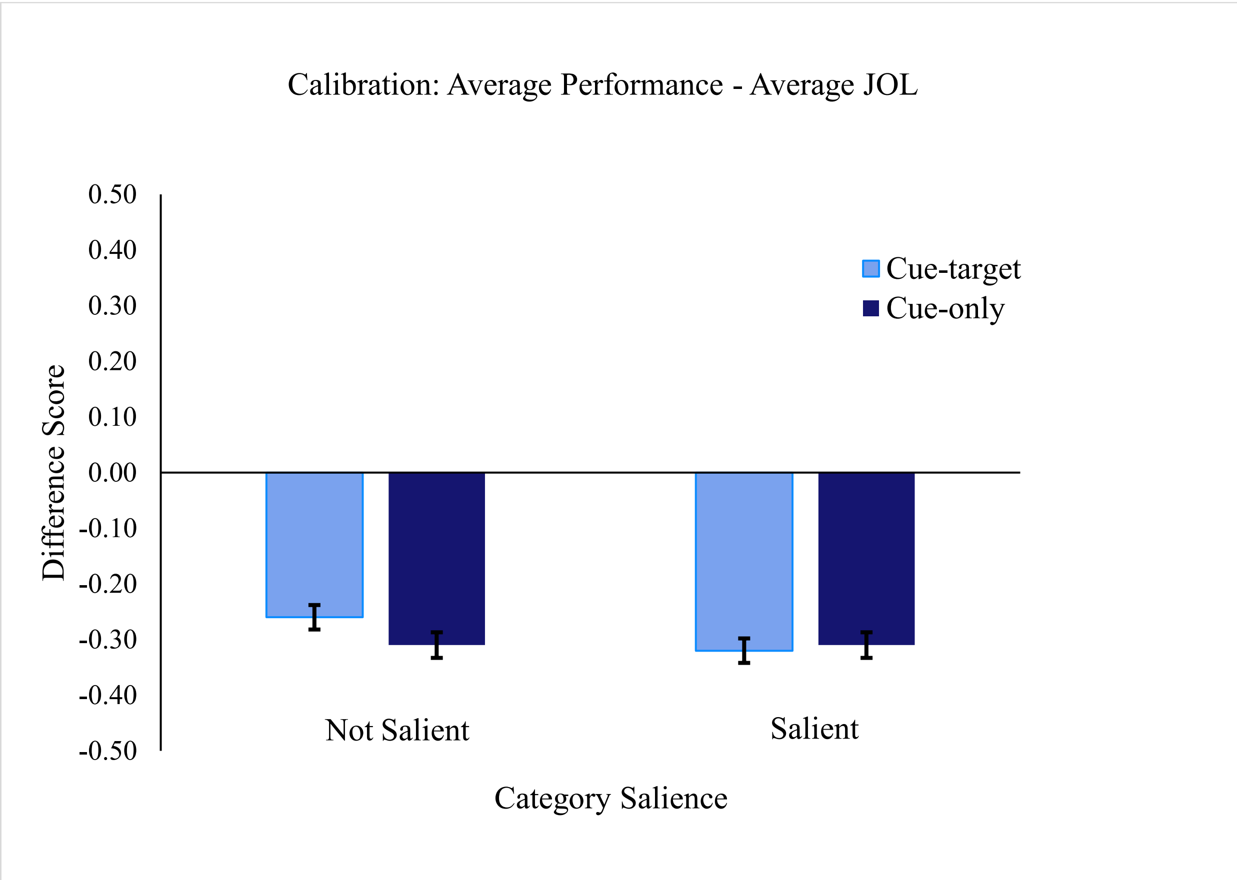
Calibration on Section A

	M	*SEM*
Cue-only		
S	-.28	.03
NS	-.22	.03
Cue-target		
S	-.24	.03
NS	-.18	.03

[a]Measured by calculating the difference between average performance on each of the 6 artists from Section A and the JOL made for each artist. The average of differences was then computed to obtain a difference score.

Figure 7

Calibration on Section A



Calibration was measured by calculating the difference between participants' average performance and their average JOLs. Error bars represent standard error.

Chapter 5: Discussion

The goal of the current study was to determine under which circumstances JOLs can facilitate inductive learning. More specifically, the goal was to disambiguate the effects of retrieval and type of processing on facilitating inductive learning, as prior literature has not yet isolated these variables. To parse the effects, I investigated whether cue-only versus cue-target JOLs and differences in category salience (salient vs. not salient) on Section A material could enhance inductive learning for Section B material. Participants studied one set of artists, made JOLs on those artists, and subsequently, studied a new set of artists before taking a final recognition test that consisted of new paintings from the artists in both sets.

On average, inductive learning improved from Section A to Section B, which is consistent with findings from Lee and Ha (2019). Although not statistically significant, participants who made cue-only JOLs performed better on the second set of material than those who made cue-target JOLs. This result suggests that when only the cue is present, JOLs may facilitate inductive learning. One reason for this finding could be that making cue-only JOLs prompted participants to engage in a covert retrieval attempt by trying to retrieve the target (viz. the artist's name). Engaging in retrieval, particularly between studying different sets of material, can reduce mind wandering while studying the subsequent set of information; this reduction in mind wandering increases attention and enhances learning (Szpunar et al., 2014). Yan and Schacter (2018) also found that retrieving previously studied categorical material can facilitate learning of subsequent new material, demonstrating a forward effect of testing. In the present study, a covert retrieval attempt may have facilitated learning of Section B, creating a forward effect of JOLs. However, because we cannot definitively conclude that participants engaged in a covert retrieval attempt, future research could investigate whether JOLs that require an overt retrieval attempt can facilitate inductive learning.

Importantly, there was a marginal interaction in the predicted direction, suggesting that the extent to which the type of JOLs facilitate inductive learning may be dependent on category salience. Specifically, cue-only JOLs only enhanced learning when the category was salient as opposed to not salient. Prior work has demonstrated that retrieval can assist learners with differentiating between categories. Thus, having a retrieval opportunity when the category was salient may have allowed learners to pick up differences between artists' painting styles, allowing them to abstract a rule (Szpunar et al., 2008; Yang et al., 2018). In a similar vein, the sequential attention theory posits that learners sometimes compare current exemplars to previously studied exemplars when learning categories (Carvalho & Goldstone, 2015). Thus, making the category salient may have allowed learners to compare members of a shared category, which has been demonstrated to be an effective method for inductive learning (Lowenstein, 2010).

The advantage of category-salient, cue-only JOLs may also stem from learners' shift in encoding processes from item-level to relational processing. This shift may have enhanced learners' ability to attend to and process the similarities between members of the same category, which allowed them to discover the deeper relational structure between the exemplars (Goldwater et al., 2018; Rottman et al., 2012). On the other hand, when the category was not salient, learners may have continued to encode exemplars at the item-level, which may have masked the relationship between paintings from the same artists and thus, the artists' painting styles. The latter theory is consistent with the current findings, as well as findings from Lee and Ha (2019). That is, when participants were encouraged to focus on individual examples rather than relational information, JOLs did not enhance learning. Together, these findings suggest that JOLs may only facilitate inductive learning when participants engage in covert retrieval and are focused on the category as a whole, rather than an individual example of the category.

One limitation of interpreting this finding is that participants in the category salient condition were given three paintings from each artist instead of one painting. In the cue-only category salient condition, participants may have actually engaged in three separate covert retrieval attempts (one for each painting). Therefore, the benefit of cue-only category salient JOLs over cue-only not salient JOLs may stem entirely from more retrieval attempts rather than from differences in encoding processes.

One question that emerges from these results is why category salience alone did not facilitate inductive learning. Some literature suggests that individual differences could have contributed to this finding. For example, some individuals are more prone to using rule-abstraction as a learning strategy, while others tend to rely on memorizing specific examples (Little & McDaniel, 2015a). In this vein, Goldwater and colleagues (2018) found that when learning via rule-abstraction, having both a task that highlights relationships among categories (i.e., category salience) and instructions that encourage rule-abstraction are sometimes necessary for inductive learning to occur. Therefore, category salience, alone, may not be enough to teach students how to learn inductively. The addition of a retrieval attempt may have been enough to key participants to abstract a rule in a similar vein as explicit instructions. Future studies should investigate how individual differences in study strategies may influence the effectiveness of using different types of JOLs to facilitate inductive learning.

Metacognitive accuracy was examined in two ways: resolution and calibration. Resolution, the ability to correctly determine which items or categories have been learned better than others (Dunlosky & Metcalfe, 2009; Jacoby et al., 2010), was low for all conditions. Contrary to my hypotheses, when the category was salient, cue-target JOLs actually led to numerically better resolution than cue-only JOLs. In contrast, when the category was not salient, cue-only JOLs led to better resolution than cue-target JOLs. However, these comparisons were not significant after running Bonferroni-corrected post-hoc tests and because they were not in the predicted direction, these patterns should be interpreted cautiously. Moreover, this finding is inconsistent with prior literature that states learners are usually able to determine which categories they have learned better than others (Jacoby et al., 2010; Lee & Ha, 2019). Although prior literature also suggests that cue-only JOLs lead to better metacognitive accuracy (Rhodes & Tauber, 201l; Rhodes, 2016), participants in the cue-only condition may have retrieved the wrong artists' name when making their JOL and because feedback was not provided, would not have necessarily known that their attempt was incorrect nor become aware that there was a gap in their knowledge (Metcalfe, 2009).

Calibration, or how well one can accurately predict one's overall performance on a task, was also low for most participants. In contrast with my hypotheses, participants were generally overconfident and inaccurate in their ability to predict their overall performance. Overconfidence, particularly in the category salient and cue-target conditions, may have been a result of perceptual fluency. That is, participants may have relied on how accessible exemplars of each category were at the time of the JOL, leading to overconfident judgments (Dunlosky & Tauber, 2016; Koriat & Bjork, 2005; Rhodes & Castel, 2008). These findings could be problematic in practice, as prior research has demonstrated that learners use their metacognitive assessments to inform their study strategies (Metcalfe, 2009; Son & Metcalfe, 2000). Recent literature suggests that individuals with poorer task performance may be worse at accurately assessing what they do or do not know (Vourre & Metcalfe, 2021). Thus, because performance was at floor level, it may be unsurprising that metacognitive accuracy was low. One caveat of this explanation is that tasks that permit guessing, such as the multiple-choice test used in the present study, sometimes lead to artificially enhanced metacognitive accuracy (Vourre & Metcalfe, 2021).

In conclusion, cue-only JOLs may show promise in facilitating inductive learning when the category is salient. When the category is not salient, JOLs likely provide no advantage in fostering inductive learning. However, it is unknown whether these results, albeit marginal, persist in other inductive learning tasks (e.g., biological taxonomy or motor skills). It may be that the facilitative effect of JOLs on inductive learning depends on the task at hand (Kattner et al., 2016). To my knowledge, this study is the first to directly investigate the role of category salience on metacognitive accuracy. Future studies should take a closer look at the effect of category salience on metacognitive accuracy to determine whether there are circumstances in which category salience can improve metacognitive accuracy. Additionally, performance was at floor level, which may have made it difficult to see any significant effects. Therefore, another future direction could be to make the study time self-paced or use easier materials in an attempt to increase overall performance. The findings from this study add to the body of literature on inductive learning and metacognition, both of which could be important when considering practical implications, particularly in academia.

References

Bennett, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning and Verbal Behavior*, *14*, 123-144.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin, 145*(11), 1029-1052.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1118-1133.

Butler, A. C., Black-Maier, A. C., Raley, N. D., & Marsh, E. J. (2017). Retrieving and applying knowledge to different examples promotes transfer of learning. *Journal of Experimental Psychology: Applied, 23*(4), 433-446.

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory and Cognition, 43*, 193-205.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563-1569.

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*, 281-288.

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. http://dx.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.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory and Cognition, 33*, 1096-1115.

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 and Cognition, 20*(4), 374-380.

Dunlosky, J., & Tauber, S. K. (2016). *The Oxford Handbook of Metamemory*. Oxford University Press.

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

Goldwater, M. G., Don, H. J., Krusche, M. J. F., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General, 147*(1), 1-35.

Goldwater, M. B., & Genter, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137-153.

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), 1441-1451.

Jönsson, F. U., Hedner, M., & Olsson, M. J. (2012). The testing effect as a function of explicit testing instructions and judgments of learning. *Experimental Psychology, 59*(5), 251-257.

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17*(4), 471-479.

Kattner, F., Cox, C. R., & Green, C. S. (2016). Transfer in rule-based category learning depends on the training task. *PLOS One*, *11*(10), 1-17.

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory and Cognition, 31*(6), 918-929.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory and Cognition*, *36*(2), 416-428.

Koriat, A. (2019). Confidence judgments: The monitoring of object-level and same-level performance. *Metacognition and Learning, 14*, 463-478.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187-194.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*(2), 219-224.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585-592.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989-998.

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.

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. http://dx.doi.org/10.1037/edu0000339

Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory and Cognition, 43*, 283-297.

Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory and Cognition*, *43*, 85-98.

Loewenstein, J. (2010). How one's hook is baited matters for catching an analogy. *Psychology of Learning and Motivation, 53,* 149-182.

McDaniel, M. A., Cahill, M. G., & Bugg, J. M. (2015). The curious case of orthographic distinctiveness: Disruption of categorical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 104-113.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20*(4), 516-522.

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effect of retrieval processing. *Memory and Cognition, 17*, 423-434.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159-163.

Metcalfe, J., & Finn, B. (2008a). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174-179.

Metcalfe, J., & Finn, B. (2008b). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1084-1097.

Middlebrooks, C. D., & Castel, A. D. (2018). Self-regulated learning of important information under sequential and simultaneous encoding conditions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(5), 779-792. http://dx.doi.org/10.1037/xlm0000480.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*(4), 267-270.

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 UP.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615-626.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgements of learning on metacognition: A meta-analytic review. *Psychological Bulletin*, *137*(1), 131-148.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255.

Rottman, B. M., Gentner D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science, 36,* 919-932.

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.

Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language, 51*(1), 71-79.

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.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3*(5), 315-316.

Storm, B. C., & Bjork, R. A. (2010). Optimizing retrieval as learning a new event: When and why expanding retrieval practice enhances long-term retention. *Memory and Cognition, 38*(2), 244-253.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, *3*(3), 161-164.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 6313-6317. http://dx.doi.org/10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1392-1399. http://dx.doi.org/10.1037/a0013082

Tekin, E., & Roediger, H. L., III. (2021). The effect of delayed judgments of learning on retention. *Metacognition and Learning.* https://doi.org/10.1007/s11409-021-09260-0

Vourre, M., & Metcalfe, J. (2021). Measures of relative metacognitive accuracy are confounded with task performance in tasks that permit guessing. *Metacognition and Learning.* https://doi.org/10.1007/s11409-020-09257-1

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), 485-492.