Memory & Cognition (2022) 50:512–526
https://doi.org/10.3758/s13421-021-01207-9

Effects of category learning strategies on recognition memory

Kevin O'Neill[1,2] · Audrey Liu[1,2] · Siyuan Yin[1,3] · Timothy Brady[4] · Felipe De Brigard[1,2,5,6]

Accepted: 7 July 2021 / Published online: 19 July 2021

Abstract

Extant research has shown that previously acquired categorical knowledge affects recognition memory, and that differences in category learning strategies impact classification accuracy. However, it is unknown whether different learning strategies also have downstream effects on subsequent recognition memory. The present study investigates the effect of two unidimensional rule-based category learning strategies – learning (a) with or without explicit instruction, and (b) with or without supervision – on subsequent recognition memory. Our findings suggest that acquiring categorical knowledge increased both hits (Experiments 1 and 2) and false-alarms (Experiment 1) for category-congruent items regardless of the particular strategy employed in initially learning these categories. There were, however, small processing speed advantages during recognition memory for both explicit instruction and supervised practice relative to neither (Experiment 2). We discuss these findings in the context of how prior knowledge influences recognition memory, and in relation to similar findings showing schematic effects on episodic memory.

Keywords Category learning · Recognition memory · Supervised learning · Schema

Introduction

At least since Bartlett's pioneering work, it has been known that episodic memory is influenced by previously acquired knowledge (Bartlett, 1932). Bartlett suggested that each individual experience is encoded, not only as an individual event, but also as related to a knowledge structure of similar previously encoded experiences. He called these knowledge structures *schemas*, and the notion has remained in the memory literature ever since. While studies on the effects of schematic knowledge on episodic memory vary, they have produced two consistent sets of results. On the one hand, participants are more likely to remember schema-inconsistent relative to schema-consistent items. For instance, after having read passages depicting stereotypical activities (e.g., going to a restaurant), participants are more likely to remember abnormal or unusual situations in an otherwise typical story relative to normal or usual occurrences (Bower, Black, & Turner, 1979; for a review, see Rojahn & Pettigrew, 1992). On the other hand, participants are more likely to false alarm to schema-consistent relative to schema-inconsistent lures. In a classic study, Brewer and Treyens (1981) asked participants to wait at a carefully staged office for 35 s, after which participants were transported to a different room where they were asked to recall the items in the office they were just at. Their results show that participants were more likely to falsely recall lure items one would have normally or usually found in an office, relative to items that were abnormal or unusual (see also Lampinen, Copeland, & Neuschatz, 2001).

Because schema acquisition takes time, and the resultant schemas are likely to be quite complex (e.g., we all have a vast amount of semantic knowledge about what occurs or doesn't occur at doctor's offices; Bower et al. 1979), research on schematic influences on recognition memory often capitalizes on participants' pre-acquired schemas. As a result, most studies usually employ one of two experimental strategies: either within-subject designs to evaluate effects of pre-acquired schemas on different recognition tests (Graesser & Nakamura, 1982; Roediger & McDermott, 1995), or

✉ Kevin O'Neill
kevin.oneill@duke.edu

1 Center for Cognitive Neuroscience, Duke University, Durham, NC 27708, USA

2 Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

3 Department of Marketing, University of Pennsylvania, Philadelphia, PA, USA

4 Department of Psychology, University of California in San Diego, San Diego, CA, USA

5 Department of Philosophy, Duke University, Durham, NC, USA

6 Duke Institute for Brain Sciences, Duke University, Durham, NC, USA

between-subject designs where participants with different pre-acquired schemas face identical recognition tests (Castel et al., 2007). Unfortunately, neither of these strategies directly manipulates schema acquisition, leaving thus unanswered questions as to how precisely schematic structures come to affect recognition memory later on.

There is another knowledge structure that has been shown to affect episodic memory, namely *categories*. Most work on categorical knowledge focuses on how it influences perceptual classification and discrimination (Ashby & Maddox, 2005), as well as how general background knowledge influences the acquisition of new categorical information (Murphy & Allopenna, 1994; Heit, 1998). However, there are also a few studies that have directly investigated how acquiring new categorical knowledge influences subsequent recognition. For example, Palmeri and Nosofsky (1995) instructed participants to learn to categorize geometric shapes according to a simple rule. While most stimuli were rule-consistent, there were some exceptions (i.e., rule-inconsistent items). After learning, participants completed a recognition test that involved old as well as new items that were either rule-consistent or rule-inconsistent. Their results suggest that participants had much better recognition of the rule-violating items relative to rule-conforming ones. In a related study, Sakamoto and Love (2004) manipulated the strength of the category rule by varying the number of rule-conforming and rule-violating items. When the rule was stronger (i.e., included fewer rule-inconsistent items during learning), exceptions were remembered better than when the rule was weaker.

Despite these apparent similarities in the influence of schemas and categories on memory, research on schemas has proceeded largely separated from research on categorical knowledge. There are several reasons behind this historical division. For one, researchers disagree about the precise characterization of both notions. Just as there are several theories about the nature of categories (Smith & Medin, 2013) as well as the psychological processes underlying category learning (Love, 2013), there are equally numerous views on the nature of schemas and their acquisition (Ghosh & Gilboa, 2014). There are also differences in research goals. While most research on categorization has focused on the processes by means of which we come to acquire categorical knowledge (Ashby & Maddox, 2005), research on schemas has mainly focused on the effects of already acquired schemas on other cognitive processes, such as emotion, memory, and decision-making. As a result, researchers in both fields have pursued different questions, and have employed distinct experimental paradigms and analytic strategies.

This was not always the case. In a classic study on categorization, Posner and Keele (1968) trained participants to learn patterns of dots that varied in terms of their distortion from a prototype, which they explicitly equated to Bartlett's schema (see also Attneave, 1957). Following training, participants were shown old dot patterns – i.e., patterns they have seen during learning – as well as new patterns, which were either distorted variants of the previously seen patterns or the unseen prototypes underlying the patterns learned during training – which Posner and Keele called "the schemas of the memorized instances" (p. 354). They found that participants had strong memories for the old patterns and strong memories even for the unseen prototype, relative to previously unseen distorted patterns. Likewise, their reaction times were equivalent for old and prototypical patterns, but longer for new ones. This result constituted one of the first demonstrations that prototypicality affected recognition and reaction times in a related memory task, and also one of the first studies linking the concept of memory schema with the notion of prototype within the context of category learning.

In recent years, a handful of researchers have sought to explicitly revive the idea that there are important connections between schematic and categorical knowledge in the context of recognition memory (Clapper, 2008; Davis et al., 2014; Love, 2013; Sakamoto, 2012). For instance, both Sakamoto (2012) and Love (2013; see also Sakamoto & Love, 2003, 2004) have argued that if one understands schematic learning as a process that involves the gradual acquisition of expectations based on prior experience, then one can think of category learning as a tantamount exercise, whereby one builds "schema-like representations in which rule-following items are encoded as a set of expectations, and rule-violating items are stored separately" (Sakamoto, 2012: 2961). Under this understanding of schema, they hypothesized that the memory advantages for exceptional items (i.e., rule-violating and schema-inconsistent) as well as the increased rate of false alarms to rule-conforming and schema-consistent lures, may derive from similar cognitive and neural mechanisms (Davis, Love, & Preston, 2012).

Inspired by this way of thinking about schematic and categorical knowledge, and seeking to further contribute to the integration of these two perspectives within the context of recognition memory, De Brigard et al. (2017) developed a paradigm to explore how learning a novel category influences subsequent memory for items belonging to the learned category relative to items that belonged to a different not-learned category or to no category at all. As mentioned, one of the key differences between research on schematic versus categorical influences on recognition memory is that the acquisition of schematic knowledge is typically assumed and rarely, if ever, manipulated. By contrast, in category-learning paradigms, the acquisition of categorical knowledge is usually well controlled and manipulated, whereas the memory component is rarely, if ever, separated from the learning stage. As such, De Brigard et al. (2017) made use of a category learning manipulation to garner better control over the acquisition of the knowledge structure and then, in separate stages of the paradigm, evaluate its impact on the subsequent

encoding and retrieval of related material. More precisely, in the *learning* stage, participants learned to categorize computer-generated flowers according to a simple unidimensional rule – a single feature (e.g., yellow petals) – which was counterbalanced across participants. These flowers constituted the *learned* category. Another feature, which occurred equally as frequently as the criterion feature for the learned category, was also counterbalanced across participants and overlapped with the learned category half the time. Flowers that included this second feature belonged to the *not-learned* category. Finally, some flowers belonged to *both* categories, and others belonged to *neither* category. In the *study* stage, participants studied previously unseen items from the learned category, the not-learned category, both categories, and neither category. Finally, at the *testing* stage, participants saw old and new items from the learned category, the not-learned category, both categories, or neither category. Across several experiments, De Brigard et al. (2017) found that learning a category increased both hits and false alarms for category-consistent stimuli in a subsequent memory test, whereas no such effect was present for the equally frequently presented yet not learned category. More recently, employing a version of the same paradigm, Yin et al. (2019) replicated these findings and found that participants who learned the category better showed improved old-new discrimination in a recognition test, relative to those who learned the category less well, suggesting that expertise in category learning enhanced memory performance.

Together, these findings offer strong evidence that learning a new category structure prior to studying items for a recognition test can influence recognition memory in a manner that is consistent with reported results in the schema literature. However, since De Brigard et al. (2017) only tested categories learned with supervised practice in the absence of instruction, it is unclear whether these results generalize to other forms of learning. In particular, it is possible that learned categories do not exhibit schema-like influences on recognition memory when those categories are learned without feedback or through explicit instruction of the category rule. No study to date has examined whether such differences in how categories are learned impacts subsequent memory, provided that similar classification accuracy is achieved.

There are two reasons why we might expect to find effects of learning strategy. First, learning strategies are thought to determine a category's underlying representation in distinct memory systems, which has been shown to affect categorization accuracy (Ashby, Maddox, & Bohil, 2002; Chandrasekaran et al., 2016; Ruge & Wolfensteller, 2010; Sakamoto & Love, 2010). For instance, compared to supervised learning, unsupervised learning has been associated with decreased categorization performance (Ashby, Maddox, & Bohil, 2002; Edmunds, Milton, & Wills, 2015), greater sensitivity to feature saliency, feature variability, and inter-feature correlations (Bröker, Love, & Dayan, 2021; Hsu & Griffiths, 2010; Levering & Kurtz, 2015), and a strong preference for linear over non-linear categories (Love, 2002). Other studies have investigated the impact of explicit instruction of the category rule: Allen and Brooks (1991), for instance, found increased classification accuracy but reduced speed for categories learned with explicit instruction relative to categories learned without instruction when rule congruency conflicted with similarity information. They took this result to mean that information learned through instruction and through practice interact; that is, practicing a learned semantic rule does not simply automatize the rule; rather, it associates the rule with the particular items and episodic contexts encountered during practice.

Another reason that learning strategies might influence memory is that feedback during learning (Dickerson & Adcock, 2018; Mather & Schoeke, 2011; Seabrooke et al., 2019), and even the act of choosing responses alone (Leotti & Delgado, 2011), is associated with increased motivation. Though this effect is reduced under the presence of instructed learning (Li, Delgado, & Phelps, 2011), this increased motivation has been associated with enhanced memory (Murty, DuBrow, & Davachi, 2015; Potts, Davies, & Shanks, 2019). As a result, if people are more motivated to successfully classify items as belonging to a particular category, they may be more likely to remember them later on.

For these two reasons, we predicted that learning a category with instruction and/or practice should impact subsequent memory, for instance, by enhancing schema-like effects (i.e., exhibiting a larger increase in hits/false alarms (FAs) for category-consistent items) and by decreasing reaction times during recognition. The current study explores this question by modifying De Brigard et al.'s (2017) paradigm to investigate the effects of four category learning strategies on recognition memory: (a) learning with and without explicit instruction, and (b) learning with and without supervised practice.

Experiment 1

By comparing recognition memory for stimuli in a learned category versus stimuli that belonged to a not learned or to neither category, De Brigard et al. (2017) showed increased hit and FA rates to lures from the learned category, but not so for lures from either the not learned or the neither category. In the current experiment, we tested whether such memory effects were modulated by the conditions under which the category was learned. Specifically, we were interested in whether memory accuracy and response time varied (a) when participants were explicitly instructed of the category rule or had to learn the rule for themselves, and (b) when participants actively practiced categorizing stimuli with supervised feedback or simply witnessed those stimuli being categorized.

Method

Participants To match the statistical power obtained in De Brigard et al. (2017) in each between-subjects condition, 867 participants were recruited via Amazon Mechanical Turk (https://www.mturk.com). All participants were from the USA, had at least 100 approved HITs, had an overall HIT approval rate of at least 95%, and received $2.00 in compensation. As we were interested in how successfully learned categories impact memory performance, data from 134 participants were excluded because of failure to learn the category above 85% accuracy during the last 20 trials of learning, as in De Brigard et al. (2017), leaving 733 participants (151 Practiced only, 208 Instructed only, 184 Both, 190 Neither) for data analysis. Out of 39,582 test-phase trials across all participants, 188 trials (0.48%) with response time greater than 3 standard deviations (SDs) from the mean (i.e., above 15.16 s) were also discarded. All participants provided informed consent in accordance with Duke University Institutional Review Board (IRB).

Materials Stimuli consisted of MATLAB (2018b)-generated flowers, used previously in De Brigard et al. (2017). These flowers varied over five features (i.e., petal number, petal color, center shape, center color, and sepal number), with each feature taking three possible values (Fig. 1A). Flowers were displayed on the center of an otherwise white screen.

Procedure The procedure, which closely followed Experiment 4 of De Brigard et al. (2017), included three phases: learning, study, and test. The experiment began with an instruction screen (~30 s) detailing the five stimulus features and their possible values, with two example stimuli displayed for illustration. Participants were instructed that they would see flowers on the screen, one at a time, and would be asked to determine whether each flower belonged to the species *avlonia*. The feature and value that constituted this Learned category (*avlonia*) was counterbalanced across participants. To isolate effects of the Learned category from possible effects of non-conceptual stimulus features, participants were also assigned a Not-Learned category, randomly defined by a value of a different feature, of which they were unaware. The Not-Learned category was never mentioned to the participants, was statistically independent of the Learned category, and was counterbalanced across participants, serving only as a baseline for analysis. During each of the three phases of the experiment, each value of each feature was displayed in one-third of the trials for that phase, so that the co-occurrence of all feature/value combinations was uniform. Accordingly, one-third of all flowers presented were members of the Learned category.

Additionally, we introduced two counterbalanced between-subjects manipulations on learning: whether the participant was explicitly instructed of the Learned category's rule (Instruction: Instructed, Not-Instructed), and whether the participant actively categorized flowers during learning or merely watched as flowers were categorized on the screen (Practice: Practiced, Not-Practiced). In the Instructed condition, participants were told how to identify *avlonias* (e.g., "*Avlonias* are flowers with six petals"). In the Not-Instructed condition, participants were told only that they would have to learn what feature and value defined the species *avlonia*. In the Practiced condition, participants completed 72 self-paced trials in which they pressed the "Y" key if the flower was an *avlonia*, or the "N" key otherwise. Immediate feedback ("Correct"/"Incorrect") was presented after each key-press for 1 s. In the Not-Practiced condition, participants instead passively viewed 72 trials in which a flower was shown for 3 s, and a categorization ("Avlonia"/"Not Avlonia") was presented immediately after for 1 s. Of the 72 flowers presented, 16 flowers were in the Learned category but not the Not-Learned category, 16 flowers were in the Not-Learned category but not the Learned category, eight flowers were in both categories, and 32 flowers were in neither category.

In the study phase, participants read instructions (for a minimum of 30 s) in which they were asked to memorize 18 flowers. Each flower was shown for 5 s after a 1-s inter-trial interval. None of these flowers were shown previously. Of these 18 flowers, four were in the Learned category but not the Not-Learned category, four were in the Not-Learned category but not the Learned category, two were in Both categories, and eight were in Neither category. Participants were told that they would receive a bonus if they could remember a high number (85%) of flowers.

Finally, in the test phase, participants read instructions (for a minimum of 30 s), in which they were told that they would see 54 flowers, one by one, and asked to press the "Y" key if the flower was old, or "N" otherwise. Each trial was self-paced with a 1-s inter-trial interval. Of these 54 flowers, 18 were presented during study. Of the remaining 36 flowers (lures), eight were in the Learned category but not the Not-Learned category, eight were in the Not-Learned category but not the Learned category, four were in Both categories, and 16 were in Neither category. None of the lures appeared in the learning or study phases.

Results

Learning phase Because the participants in the Not-Practiced condition did not make responses during learning, we report results from those in the Practiced condition only. As found in De Brigard et al. (2017), participants in the Not-Instructed condition started at near chance (M = 65.4%, SD = 20.2%) categorization accuracy in the first ten trials, and gradually rose to near ceiling (M = 98.7%, SD = 3.8%) accuracy in the last ten trials. In contrast, participants in the Instructed

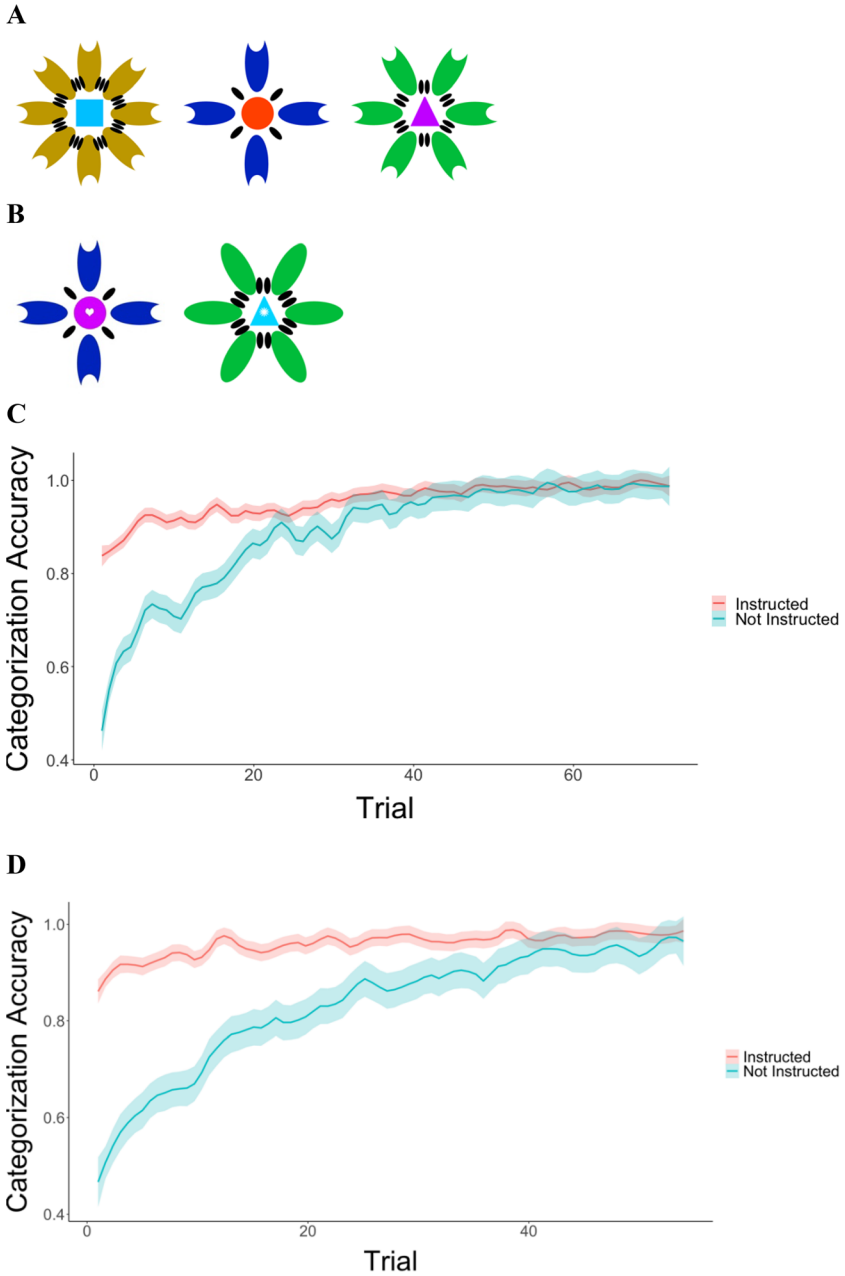

Fig. 1 (**A**) Examples of stimuli for Experiment 1. Stimuli consisted of flowers varying across five dimensions, with each dimension taking on one of three possible values: number of petals (four, six, or eight), petal color (blue, green or yellow), center shape (circle, triangle, or square), center color (orange, purple, or turquoise), and number of sepals (one, two, or three). Figure 1A depicts all possible values of the five features. (**B**) Examples of stimuli for Experiment 2. Stimuli consisted of flowers varying across seven dimensions, with each dimension taking on one of two possible values: number of petals (four or six), petal color (blue or green), center shape (circle or triangle), center color (purple, or turquoise), number of sepals (one or two), center hole shape (heart or star), and petal shape (concave or round). Figure 1B depicts all possible values of the seven features. (**C**) Learning performance for Experiment 1. Mean categorization performance for the Practice + Instruction and Practice + No-Instruction conditions. Shaded areas represent standard errors from the mean. (**D**) Learning performance for Experiment 2. Mean categorization performance for the Practice + Instruction and Practice + No-Instruction conditions. Shaded areas represent standard errors from the mean

condition began with high accuracy (M = 89.2%, SD = 17.4%), and quickly rose to near ceiling (M = 99.1%, SD = 3.1%) (Fig. 1C). These results confirm that instructing participants of the category rule allowed them to quickly learn to categorize flowers correctly.

Test phase As outcome variables during the testing phase, we measured reaction time (RT), hits, and false alarms (FAs), as well as hierarchical estimates of sensitivity (d') and bias (C). For all outcome variables, unless otherwise noted, we fit Bayesian mixed-effects models in R using the *brms* package

with five chains, 1,000 iterations of burn-in, and 1,500 iterations of sampling (Bürkner, 2017). In accordance with Bayesian parameter estimation (see Kruschke & Liddell, 2018, for a review), for each effect, we report posterior medians, 95% Highest Density Intervals (*HDI*s), probability of direction (pd), and percentage of the 95% *HDI*s inside a Region of Practical Equivalence (ROPE). Similar to a *p*-value, probability of direction is the proportion of posterior samples greater than (or less than) 0 and is a metric of effect existence, where $pd > 97.5\%$ is suggestive of an effect at $\alpha = .05$ (for a discussion on these Bayesian approaches, see Makowski, Ben-Shachar, Chen, & Lüdecke, 2019). In contrast, the percentage of the 95% *HDI* inside a ROPE is a metric of effect significance that counts the proportion of posterior samples inside a null region (Kruschke, 2011). We use < 5% and > 95% as rough benchmarks for rejecting and accepting the null, respectively.

Hit and false alarm rates For hits and FAs, we used 2 (Learned Category) × 2 (Not-Learned Category) × 2 (Practiced) × 2 (Instructed) Bayesian mixed-effect models with random intercepts for each subject, uncorrelated random slopes for Learned Category and Not-Learned Category, and Bernoulli distributions with a logit link. We used weakly informative Gaussian priors on all coefficients, with means of 0 and standard deviations of 4, and a ROPE width of 0.15. There was strong evidence that flowers in the Learned category (*Mdn* = 0.65, *HDI* = [0.60, 0.70]) were recognized more frequently than flowers not in the Learned category (*Mdn* = 0.57, *HDI* = [0.54, 0.61]), $\beta = .32$, *HDI* = [.1, .54], *pd* = 99.9%, 3.9% in ROPE. In contrast, there was evidence against effects of Not-Learned Categories on recognition memory, $\beta = .18$, *HDI* = [-.03, .39], *pd* = 95%, 40.2% in ROPE, whether participants were allowed to practice during categorization, $\beta = -.08$, *HDI* = [-.29, .13], *pd* = .75.9%, 76% in ROPE, and whether they were given explicit instructions about the category, $\beta = -.08$, *HDI* = [-.27, .12], *pd* = 77.9%, 78.9% in ROPE. There was also evidence against all 2-, 3-, and 4-way interactions between these variables (all *pd*s < 87.5%, > 22.6% in ROPE). Posterior estimates of hit rate for Experiment 1 are shown in Figure 2A (for further results, see Online Supplementary Material).

For FAs, we again found weak evidence that lures in the Learned category (*Mdn* = 0.53, *HDI* = [0.48, 0.57]) were more likely to be falsely recognized than lures not in the Learned category (*Mdn* = 0.46, *HDI* = [0.43, 0.50]), $\beta = .26$, *HDI* = [.08, .43], *pd* = 99.8%, 9.2% in ROPE. Conversely, there was evidence against effects of the Not-Learned Category, $\beta = .08$, *HDI* = [-.07, .23], = 86.4%, 81.7% in ROPE, whether participants were allowed to practice categorization, $\beta = -.03$, *HDI* = [-.23, .17], *pd* = 62.3%, 88.5% in ROPE, and whether they were given explicit instructions about the category, $\beta = .11$, *HDI* = [-.07, .29], *pd* = .87.9%, 67.0% in ROPE. There was moderate to strong evidence against all other 2-, 3-, and 4-way interactions between these variables (all *pd*s < 97.7%, > 15% in ROPE). Posterior estimates of false alarm rate for Experiment 1 are shown in Fig. 2C (for further results, see Online Supplementary Material).

Discriminability (d') and response bias (C) Next, hierarchical estimates of SDT parameters (*d'*, *C*) were calculated using a 2 (Learned Category) × 2 (Not-Learned Category) × 2 (Practiced) × 2 (Instructed) × 2 (Old) Bayesian mixed-effect model with random intercepts for each subject, uncorrelated random slopes for Learned Category, Not-Learned Category, and Old/New status, and a Bernoulli distribution with a probit link. We used weakly informative Gaussian priors on all coefficients, with means of 0 and standard deviations of 2, and a ROPE range of 0.075. In this model, the intercept represents estimates of -*C*, the effect of Old represents estimates of *d'*, effects of our factors represent effects on *C*, and interactions with Old represent effects on *d'* (De Carlo, 1998). There was strong evidence for a main effect of Old/New status, suggesting that subjects had positive estimates of *d'*, $\beta = 0.27$, *HDI* = [0.18, 0.36], *pd* = 100%, 0% in ROPE. For response bias (C), there was evidence for an effect of Learned Category, $\beta = .16$, *HDI* = [.05, .26], *pd* = 99.8%, 3.8% in ROPE, suggesting that participants were more likely to respond 'Old' for flowers in the Learned category (*Mdn* = 0.53, *HDI* = [0.49, 0.57]) compared to flowers not in the Learned category (*Mdn* = 0.47, *HDI* = [0.44, 0.50]). Conversely, there was evidence against effects of practice during the categorization phase, $\beta = -.02$, *HDI* = [-.14, .09], *pd* = 65.3%, 80.9% in ROPE, explicit instructions about the category, $\beta = 07.$, 95% *HDI* = [-.04, .17], *pd* = 89.2%, 57.5% in ROPE, and Not-Learned category status, $\beta = .05$, *HDI* = [-.04, .14], *pd* = 86.4%, 70.7% in ROPE on response bias. There was also evidence against such effects on *d'*: practice status, $\beta = -.03$, *HDI* = [-.17, .11], *pd* = 65.7%, 70.4% in ROPE, explicit instruction status, $\beta = -.11$, *HDI* = [-.24, .01], *pd* = 95.6%, 27.6% in ROPE, Learned category status, $\beta = .04$, *HDI* = [-.11, .19], *pd* = 69.2%, 66.6% in ROPE, and Not-Learned category status, $\beta = .06$, *HDI* = [-.09, .20], *pd* = 77.8%, 59.0% in ROPE. Finally, there was evidence against all interactions between Learned category status, Not-Learned category status, Practice, Instruction, and Old (all *pd*s < 97.3%, > 11.2% in ROPE). Posterior estimates of *d'* and *C* for Experiment 1 are presented in Figs. 3A and 3C (for further results, see Online Supplementary Material).

Reaction time To examine effects on RT, we employed a 2 (Learned Category: Yes, No) × 2 (Not-Learned Category: Yes, No) × 2 (Practiced: Yes, No) × 2 (Instructed: Yes, No) × 2 (Old: Yes, No) Bayesian mixed-effect model with random intercepts for each subject, uncorrelated random slopes for Learned Category, Not-Learned Category, and Old,

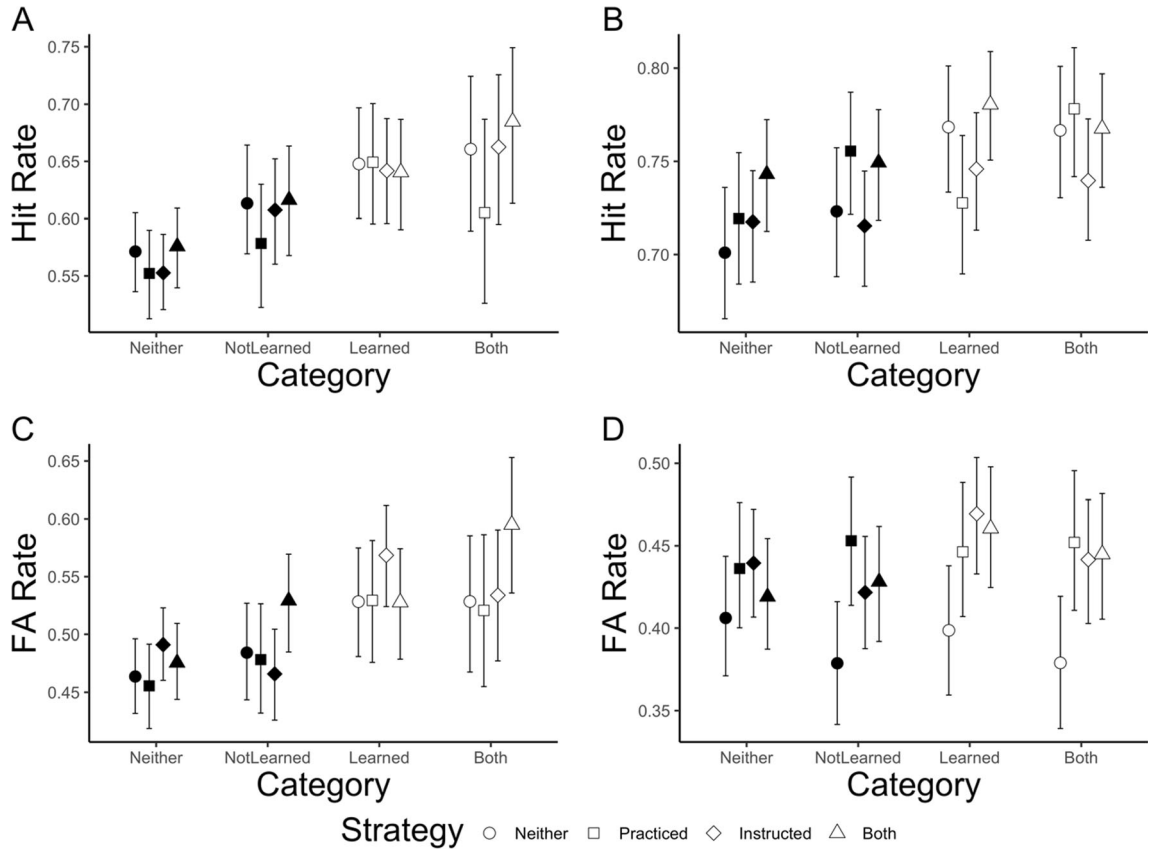

Fig. 2 Hit and false alarm (FA) rates for Experiment 1 (**A** and **C**) and Experiment 2 (**B** and **D**). The fill represents whether the stimulus was in the learned category (white = *avlonia*, black = not *avlonia*). Error bars represent 95% HDIs

predicting the parameters μ (identity link) and τ (right-skewness, log link) of an Ex-Gaussian distribution, keeping σ fixed at the population-level. The Ex-Gaussian distribution is commonly used to model RT distributions by separately accounting for the mean (μ) and skewness (τ), thereby reducing the effect of outliers on the mean in such positively skewed distributions (Balota & Yap 2011). We used weakly informative Gaussian priors on all coefficients, with means of 0 and standard deviations of 2.5 for μ, and with means of 0 and standard deviations of 1 for τ, and a ROPE width of [-.157, .157], which represents a standardized effect of 0.1. To aid convergence, we used 3,000 sampling iterations with a thinning rate of 3. Posterior medians and 95% *HDI*s on RTs are reported in Table 2A (for further results, see SI). We found evidence against significant effects and interactions of all factors on μ ($pd < 0.927$, > 76% within ROPE). While there was evidence for the existence of effects of some factors on τ ($pd < 0.99$), these effects were so small as to be negligible (> 29.1% within ROPE; see Table S7 in Online Supplementary Material). These results suggest that there were no substantial differences in RT arising from our manipulations. Posterior estimates of the mean and skewness of RTs for Experiment 1 are shown in Fig. 4A and 4C (for further results, see Online Supplementary Material).

Discussion

Experiment 1 provided initial evidence that learning a category prior to the encoding stage affects subsequent recognition memory no matter what strategy was used to learn the category. In particular, we found an increase in correct recognition (hits) and FAs for stimuli in the learned category compared to stimuli not in the learned category. Under the signal-detection framework, we characterized this effect as an increased bias to recognize stimuli in the learned category regardless of whether the stimuli were in fact old, replicating past work on category learning and memory (De Brigard et al., 2017; Yin et al., 2019). We also found that category learning had a similar impact on recognition memory irrespective of whether participants were explicitly instructed of the category rule or of whether they learned the category through supervised practice. However, this initial study has several limitations. First, although we ensured that participants who practiced categorization during learning successfully learned the category, we had no way of assessing whether the same was true for participants who did not practice in the learning phase. Additionally, although our signal-detection model was able to identify a change in response bias for stimuli in the learned category, a characterization of the full receiver operating

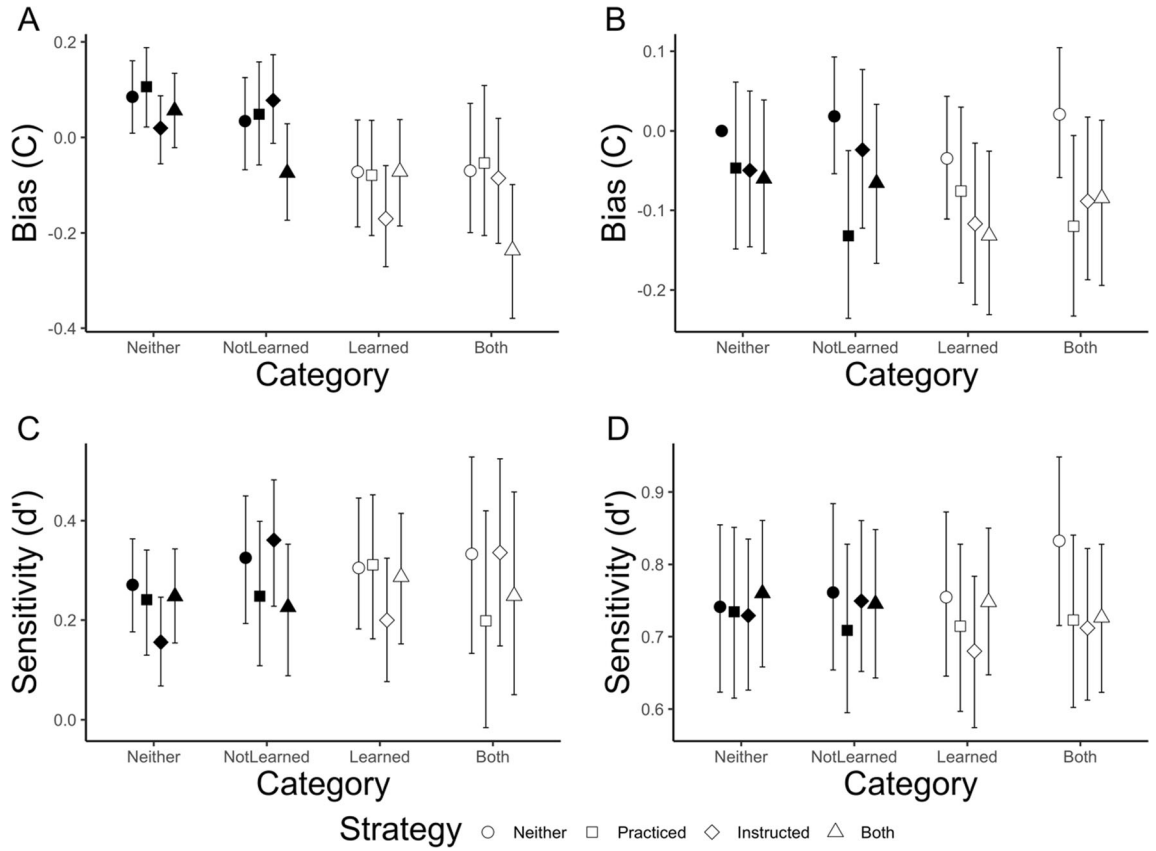

Fig. 3 Bias (**C**) and sensitivity (d') for Experiment 1 (**A** and **C**) and Experiment 2 (**B** and **D**). The fill represents whether the stimulus was in the learned category (white = *avlonia*, black = not *avlonia*). Error bars represent 95% HDIs

characteristic (ROC) curve is necessary to ensure that estimates of response bias and sensitivity are not confounded (Wixted, 2007). Finally, memory in this experiment was poor; participants had a median sensitivity of 0.27. It is possible that memory differences might arise when memory accuracy overall is increased. We sought to address these limitations directly in Experiment 2.

Experiment 2

Method

Participants To match the statistical power obtained in De Brigard et al. (2017) in each between-subjects condition, 996 participants were recruited via Prolific (https://app.prolific.co/). All participants were from the USA, had at least 100 approved HITs, had an overall HIT approval rate of at least 95%, and received $3.00 in compensation. As in Experiment 1, data from 225 participants were excluded because of failure to learn the category above 85% accuracy during the last 20 trials of learning, leaving 771 participants (167 Practiced only, 220 Instructed only, 215 Both, 169 Neither) for data analysis. Out of 49,344 test phase trials across all remaining participants, 98 trials with response time greater than 3 standard deviations (SDs) from the mean (i.e., above 31.73 s) and 41 trials with confidence rating response time greater than 15 s (9 SDs from the mean) were also discarded. All participants provided informed consent in accordance with Duke University IRB.

Materials Stimuli consisted of the same computer-generated flowers used in Experiment 1, displayed on the center of an otherwise white screen (Fig. 1B). However, to ensure that the presentation of flowers in all conditions was fully counterbalanced, we utilized flowers from Experiment 3 in De Brigard et al. (2017) with seven features (petal number, petal color, petal shape, center shape, center color, sepal number, shape of hole in center). Each feature could take two possible values (Fig. 1B).

Procedure The procedure for this experiment was exactly as in Experiment 1, with a few small changes. First, to gauge learning performance for participants in all conditions (and not just those who practiced during learning), we reduced the number of learning trials to 54 trials, and added a ten-trial learning test. The learning test had the same format as the learning phase, except that all participants (including those in the Not-

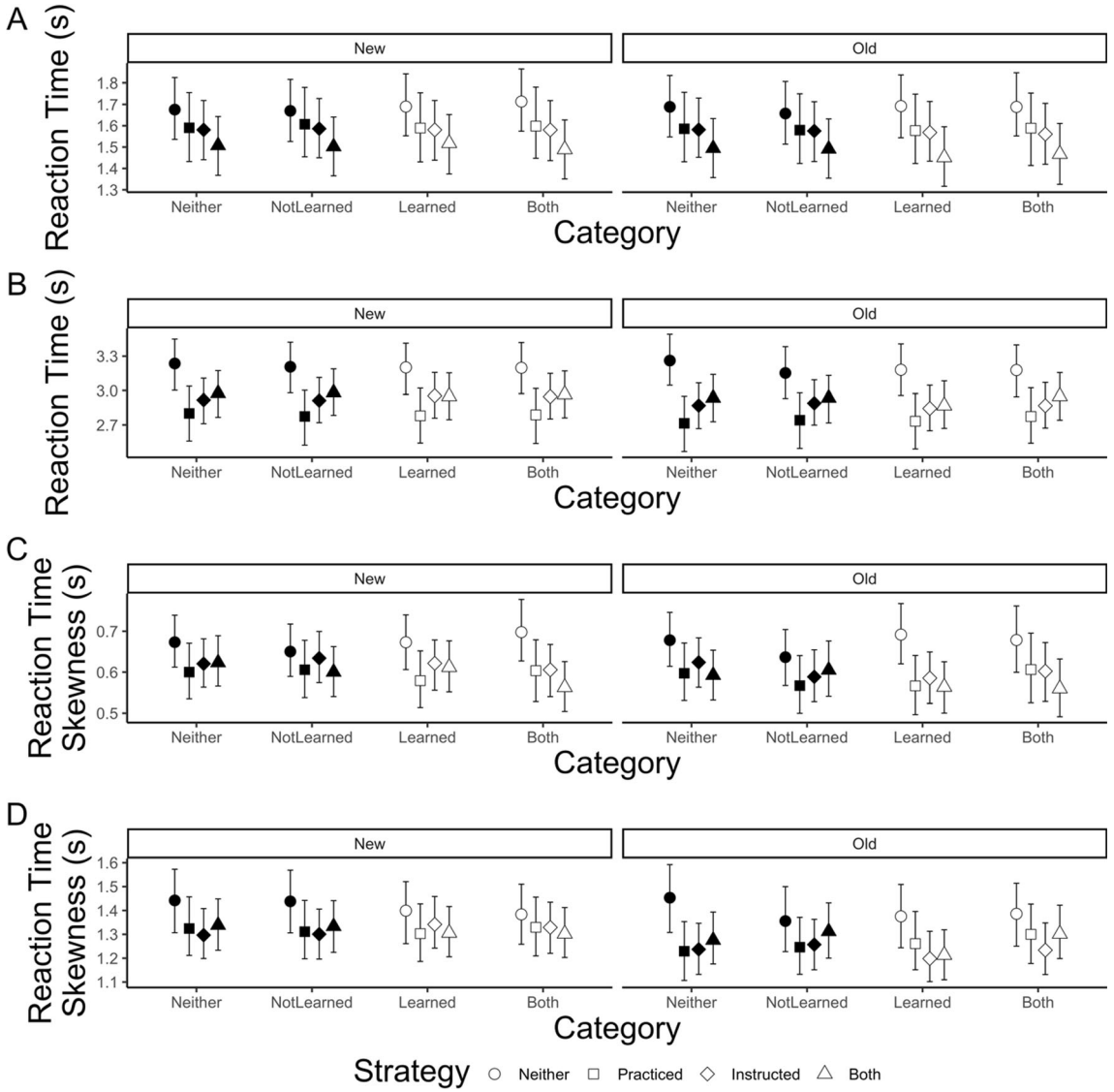

Fig. 4 Mean reaction time and skewness in reaction time for Experiment 1 (**A** and **C**) and Experiment 2 (**B** and **D**). Estimates are presented separately for new stimuli (i.e., lures) and for old stimuli. The fill represents whether the stimulus was in the learned category (white = *avlonia*, black = not *avlonia*). Error bars represent 95% HDIs

Practiced condition) actively responded whether the presented flower was an *avlonia* or not. We also increased the number of trials in the study phase to 24 flowers (six flowers in the Learned category, six in the Not-Learned category, six in Both categories, and six in Neither category), and the number of trials in the test phase to 64 (24 flowers from the study phase, ten lures in the Learned category, ten in the Not-Learned category, ten in Both categories, and ten in Neither category). To ensure that memory performance was above chance, we broke up the study and test phases into four blocks of six study trials and 16 test trials. Finally, to estimate ROC curves and de-confound response bias and sensitivity, we included a confidence rating after each trial in the test phase. Specifically, participants were asked to rate how confident they were that the presented flower was (or was not) shown in the study phase on a 3-point Likert scale (1 = not at all confident, 3 = totally confident).

Results

Learning phase As found in Experiment 1, participants in the Not-Instructed condition started at near chance categorization accuracy (M = 60.5%, SD = 20.4%) in the first ten trials, and gradually rose to near ceiling (M = 95.2%, SD = 13.2%) in the last ten trials. In contrast, participants in the Instructed condition began with high accuracy (M = 91.5%, SD = 16.5%), and quickly rose to near ceiling (M = 98.0%, SD = 7.6%) (Fig. 1D). These results confirm that instructing participants of the

category rule allowed them to immediately learn to categorize flowers correctly, but that performance was similar by the end regardless of instruction method.

Test phase As outcome variables during the testing phase, we again analyzed RT, hits, and FAs from the initial old/new response, as well as hierarchical estimates of receiver operating characteristic (ROC) curves from the confidence responses. Unless mentioned, all analyses were performed as in Experiment 1.

Hit and false alarm rates We found an effect of the Learned category on hit rate. In particular, there was strong evidence that flowers in the Learned category (*Mdn* = .76, *HDI* = [0.75 0.78]) were recognized more frequently than flowers not in the Learned category (*Mdn* = 0.73, *HDI* = [0.72, 0.74]), β = .35, *HDI* = [.15, .57], *pd* = 100%, 0.0% in ROPE. In contrast, there was evidence against effects of Not-Learned Categories on recognition memory, β = .11, *HDI* = [-.08, .31], *pd* = 86%, 67% in ROPE, whether participants were allowed to practice during categorization, β = .09, *HDI* = [-.15, .33], *pd* = 76%, 70% in ROPE, and whether they were given explicit instructions about the category, β = .08, *HDI* = [-.15, .30], *pd* = 76%, 74% in ROPE. There was also no evidence for any 2-, 3-, and 4-way interactions between these variables (all *pd*s < 98%, > 14% in ROPE). Posterior estimates of hit rate for Experiment 2 are shown in Fig. 2B (for further results, see Online Supplementary Material). In contrast to Experiment 1, we found no evidence for any substantial differences in FAs due to any of our manipulations (all *pd*s < 95%, > 34% in ROPE). Posterior estimates of the false alarm rate for Experiment 2 are shown in Fig. 2D (for further results, see Online Supplementary Material).

ROC curves We estimated hierarchical ROC curves using a 2 (Learned Category) × 2 (Not-Learned Category) × 2 (Practiced) × 2 (Instructed) × 2 (Old) Bayesian mixed-effect ordinal regression with random intercepts for each subject, uncorrelated random slopes for Learned Category, Not-Learned Category, and Old/New status, and a cumulative distribution with a probit link. Specifically, this model estimates participant's old/new responses and confidence ratings on a single ordinal scale (1 = definitely new, 6 = definitely old; see Bürkner & Vuorre, 2019, for a comprehensive tutorial on Bayesian ordinal regression). To account for the possibility of unequal variances in memory strength, we also estimated separate variances for old stimuli, with the variance for new stimuli fixed to 1 for identifiability. We used weakly informative Gaussian priors on all coefficients, with means of 0 and standard deviations of 2, and a ROPE range of 0.075. There was strong evidence that variance was lower for Old stimuli (*Mdn* = .88, *HDI* = [.80, .96]) than for new stimuli (fixed to 1), β = 0.07, *HDI* = [0.02, 0.11], *pd* = 99%, ROPE = [-.02, .02], 0% in ROPE. There was also strong evidence for a main effect of Old/New status, suggesting that subjects had positive estimates of *d'* (reported in units of standard deviations for new stimuli), β = 0.74, *HDI* = [0.62, 0.85], *pd* = 100%, 0% in ROPE. However, there was no evidence for any effects of the Learned category, the Not Learned category, Practice, or Instruction on sensitivity or bias (all *pd*s < 97%, > 30% in ROPE). Posterior estimates of *d'* and *C* for Experiment 2 are presented in Figs. 3B and D (for further results, see Online Supplementary Material). Raw ROC curves and posterior predictions of estimated ROC curves are presented in Figs. 5A and B, respectively.

Reaction time Our analysis on recognition RT is the same as in Experiment 1, with the exception that we used a ROPE of [-.26, .26], again corresponding to an effect size of .1. As in Experiment 1, there was no evidence for significant changes in the mean or skewness of RT distributions due to any of our manipulations. However, there was evidence that RT was negligibly lower when participants practiced with feedback during learning (*Mdn* = 2.85, *HDI* = [2.69, 3.00]) than when participants passively watched flowers being categorized (*Mdn* = 3.05, *HDI* = [2.91, 3.21]), β = -0.44, *HDI* = [-0.75, -0.09], *pd* = 100%, 15% in ROPE. There was also evidence that RT was negligibly lower when participants were instructed of the category rule before learning (*Mdn* = 2.92, *HDI* = [2.79, 3.07]) than when participants had to discover the rule (*Mdn* = 2.98, *HDI* = [2.82, 3.14]), β = -0.32, *HDI* = [-0.62, -0.02], *pd* = 98%, 36% in ROPE. Finally, there was evidence for a negligible interaction between Practice and Instruction, such that having both practice and instruction during learning did not further decrease RTs, β = 0.5, *HDI* = [0.06, 0.93], *pd* = 99%, 14% in ROPE. Posterior estimates of the mean and skewness of RTs for Experiment 2 are shown in Fig. 4B and 4D (for further results, see Online Supplementary Material).

Discussion

In Experiment 2 we sought to replicate Experiment 1, with a few modifications aimed at improving overall memory performance. First, we included a learning test to ensure that all participants, even those who did not make any responses in the learning phase, successfully learned the category rule. To de-confound changes in response bias and sensitivity during recognition, we recorded not just old/new judgments but also confidence judgments, allowing for estimation of full ROC curves. Finally, to increase overall memory performance, we separated the study and test phases into four blocks. Under these conditions, we replicated the finding of increased hits for items in the learned category. We also replicated the null findings that learning strategies (practice only, instruction only, both, neither) did not change the overall influence of the learned category on memory in analyses on hits, false alarms,

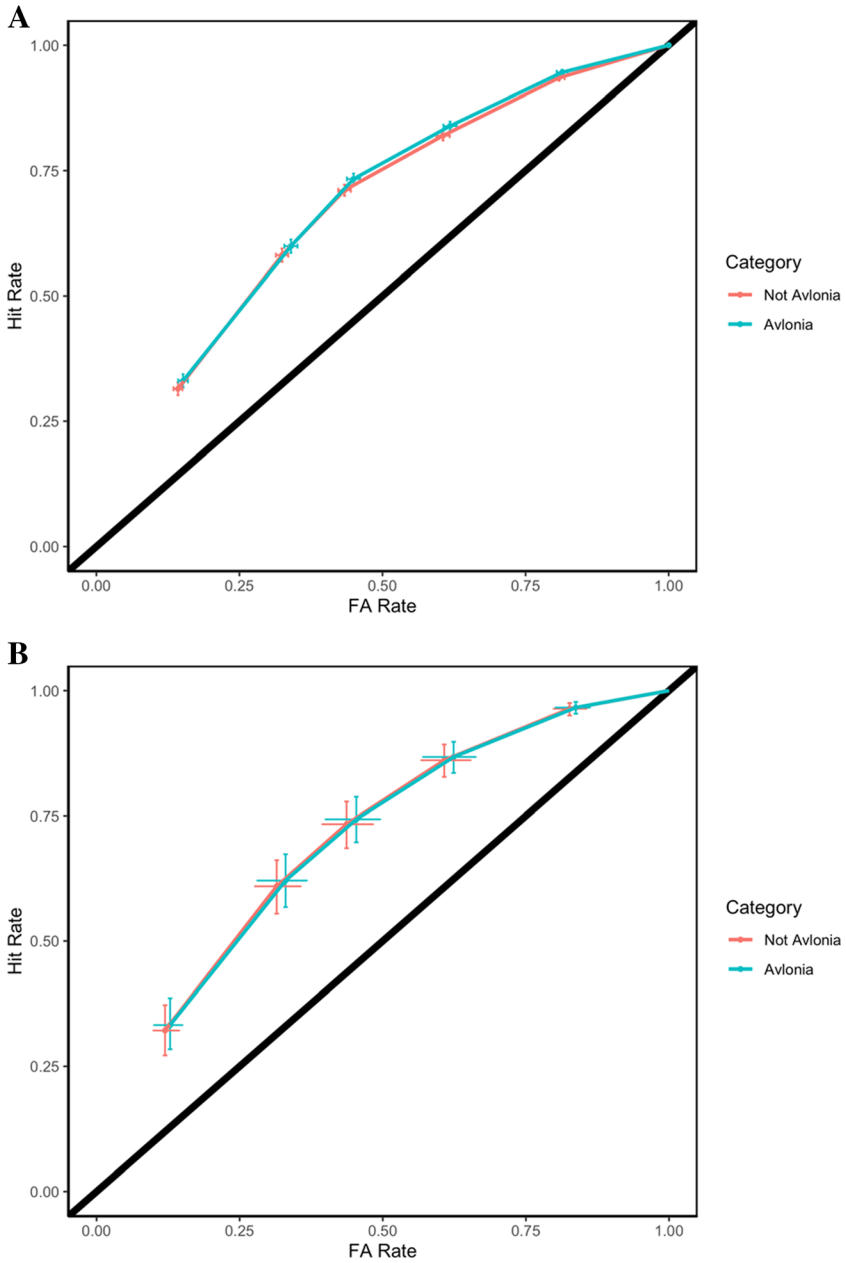

Fig. 5 Raw receiver operating characteristic (ROC) curves (**A**) and posterior estimates of ROC curves (**B**). Error bars indicate 95% CIs and 95% HDIs, respectively

and ROC curves. In contrast to these replications, however, we found no increase in FAs for items in the learned category, and further, no difference in ROCs between learned and unlearned categories. Finally, we also found that RT during recognition was slightly faster for participants who practiced or were instructed during learning.

General discussion

Ample evidence from research on schematic knowledge and on category learning has shown that previously acquired knowledge influences recognition memory. Despite common origins (e.g., Attneave, 1957; Posner & Keele, 1968), both lines of research have been largely conducted independently of one another. A consequence of this division is that investigators on each side of this divide have pursued different research questions using different experimental paradigms. On the one hand, given that schema acquisition takes time and likely varies across individuals, researchers interested in studying the effects of schemas on recognition memory tend to capitalize on pre-acquired schemas which are rarely, if ever, manipulated. On the other hand, category learning paradigms offer the possibility of tightly controlling the acquisition of

new categorical knowledge, yet they rarely, if ever, separate the learning from the study stages, which does not allow clearly differentiating the pre-acquired knowledge from the studied material.

Inspired by recent theorizing on the relationship between schematic and categorical knowledge (Love, 2013; Sakamoto, 2012), and seeking to bring the advantages of experimental controllability afforded by category learning paradigms to the schema-inspired study of how pre-acquired knowledge influences subsequent memory, De Brigard et al. (2017) developed a paradigm to explore how learning a novel category affects recognition memory for a separate set of studied items that belonged to the learned category, relative to items that belonged to a different not-learned category or to no category at all. The present study employed a variation on that paradigm to investigate whether practice and instruction during category learning have downstream effects on recognition memory. Inspired by the fact that, in the category learning literature, several findings have shown differences in classification accuracy as a function of these two category learning strategies, the experiments reported here sought to further explore whether such differences in learning strategies would also have downstream effects on recognition memory.

The current experiments yielded three main findings. First, we replicated prior results using the same paradigm (De Brigard et al., 2017; Yin et al., 2019). Specifically, we found that learning a category yielded an increase in hits (Experiments 1 and 2) and false alarms (Experiment 1) for stimuli in the learned category compared to stimuli not in the learned category. In Experiment 1 we also replicated previous findings showing a decrease in bias (i.e., *C*) for items of the learned category relative to items not in the learned category. The category learning literature offers at least two non-mutually exclusive accounts for these findings. The first one involves the role of attention during learning and encoding. Extant results indicate that during category learning, individuals tend to pay more attention to the feature(s) that best differentiate the to-be-learned-category relative to all other features (Nosofsky, 1986). This can occur by either generating an initial hypothesis that gets corroborated with every learning instance, as in the case of the instructed condition, or via the sequential iteration of hypothesis confirmation/refutation, as in the case of the non-instructed condition (Kruschke & Johansen, 1999; Mack et al., 2020; Nosofsky, Palmeri, & McKinley, 1994; Smith, Patalano, & Jonides, 1998). Since increased attention yields better encoding (De Brigard, 2012), it is likely that the attentional bias built during categorization carried over to the study stage, thus improving subsequent recognition of category-consistent items. Intriguingly, this very same attentional bias may explain the increase in false alarms found in Experiment 1. When more attention is drawn to a category-defining feature, less attention is dedicated to other features of the stimuli, which would have been critical to help to identify a test item as a category-consistent lure (Ashby & Maddox, 2005; Kruschke, 1992; Love, Medin, & Gureckis, 2004).

The second account involves processes that occur at retrieval. The influential Category Adjustment Model (CAM; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000) suggests that stimuli are encoded both as members of a category and as individuals with particular fine-grain details (Duffy et al., 2010). Both encoded representations are conceived as distributions: the former constituting an explicit prior distribution, and the latter a more-or-less noisy distribution representing the relevant encoded item. At retrieval, people integrate both levels of information to maximize accuracy. Within this framework, our results may be interpreted as suggesting that learning a category generates a prior distribution capable of biasing the estimate of individual lures as being closer to the mean of the prior (i.e., categorical) distribution, thereby increasing endorsement for category-consistent old items (Hits) as well as category-consistent new lures (FAs). Similar retrieval biases have been recently reported in studies manipulating both similarity and spatial distancing (Hemmer & Steyvers, 2009; Tompary & Thompson-Schill, 2021), further showing how CAM can be extended to understand the role of prior knowledge on recognition memory in a variety of contexts.

It is important to note, however, that the false alarm effect from Experiment 1 was not evident in Experiment 2. This may have been because study and test in Experiment 2 were broken down into smaller blocks, with testing including only 16 items per block, relative to the 54 items in Experiment 1. While breaking up the study and test section into smaller blocks definitively increased overall recognition performance, it did so at the expense of reducing false alarm rates and, along with it, the observed effect of category learning on false alarms. Whether the reduction of false alarms here is due to better attentional distribution during encoding and/or less noisy representations of the studied items at retrieval – which would imply more accurate estimates for category-consistent lures – is unclear and constitutes an interesting question for future research.

A second finding from the current study pertains to our manipulations of practice and instruction. Our analyses offered evidence against substantial differences in memory accuracy and RT between the four conditions (i.e., instructed/not-practiced, not-instructed/practiced, both, neither), suggesting that different learning strategies may be equally effective in forming stable knowledge structures in memory. More precisely, we found that neither being instructed explicitly of the category-inclusion rule nor practicing category classification during feedback-driven learning has any differential downstream effect on recognition memory. These results show that while category learning strategies may have consequences for immediate classification accuracy (Allen &

Brooks, 1991; Love, 2002), they may not differentially affect subsequent recognition memory, suggesting that these different learning strategies are equally successful in generating the categorical knowledge structure that brings about the reported increase for hit and false alarms for rule-consistent items. There may yet be conditions in which these learning strategies do induce dissociable effects on memory, however. For instance, while we presented the category label prior to presenting the image during unsupervised learning to match the conditions during supervised learning, previous work has indicated that presenting the label either simultaneously with or after the item is a more process-pure manipulation of unsupervised learning (Levering & Kurtz, 2015; Love, 2002).

An important consideration in discussing the current work is the fact that learning and motivation are profoundly intertwined, and there is plenty of evidence suggesting that motivation also impacts memory (for a recent review, see Dickerson & Adcock, 2018). As mentioned, it has been shown that choosing and receiving positive feedback are rewarding and can improve memory retrieval (Leotti & Delgado, 2011; Mather & Schoeke, 2011; Murty, DuBrow, & Davachi, 2015). But motivation is a much more complex phenomenon, and it can interact with memory in various ways. As such, we see the role of motivation during category learning and its effect on subsequent recognition as an open question, and we believe that the current paradigm may afford researchers the possibility of manipulating reward to further explore this issue experimentally.

Finally, although we have observed little difference in recognition memory between categories learned with and without supervised practice and with and without explicit instruction of the category rule, we found small decreases in RT during recognition for participants who either practiced categorization during learning or were instructed of the category rule. However, this finding was weak overall, and was only significant in Experiment 2. One potential reason that this effect was absent in Experiment 1 could be that mean RT was shorter overall in Experiment 1 (M = 1.59s, 95% HDI = [1.51, 1.67]) than in Experiment 2 (M = 2.69s, 95% HDI = [2.85, 3.06]), which reduced our ability to detect decreases in RT. Another possibility is that online data collection precluded us from obtaining precise enough measures of RT. But if replicated in future work, the preliminary finding of shorter recognition time with practice and instruction could indicate that although learning strategies may not have an impact on recognition accuracy, they may impact retrieval effort or difficulty (Mettler & Kellman, 2014; Pavlik & Anderson, 2008; Pyc & Rawson, 2009). One possibility is that different learning strategies might impact the extent to which people engage familiarity and recollection processes during episodic recognition (Atkinson & Juola, 1973; Yonelinas, Otten, Shaw, & Rugg, 2005). Another, non-incompatible possibility, is that learning strategies may modulate whether resultant memory representations are primarily reinforcement-based or structured, supported by differential activity in the striatum, or in the hippocampus and ventromedial pre-frontal cortex (Ashby & Maddox, 2005). As such, further research employing neuroimaging techniques may reveal interesting differences in the neural representations of categories associated with different learning strategies. Because this paradigm allows for testing the memory effects of newly learned categories independently of the perceptual or statistical features of the stimuli, we believe that it would be fertile ground for testing such hypotheses in future work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-021-01207-9.

References

Allen, S. W, & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology,*120, 3–19.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & cognition*, *30*(5), 666-677.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology, 56*, 149-178.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. *Attention and performance IV*, 583-612.

Attneave, F. (1957). Transfer of experience with a class-schema to identification-learning of patterns and shapes. *Journal of experimental psychology*, *54*(2), 81.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*(3), 160-166.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. : Cambridge University Press.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology,* 11, 177–220.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207–230.

Bröker, F., Love, B. C., & Dayan, P. (2021). When unsupervised training benefits category learning.

Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1 - 28.

Bürkner, P. C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77-101.

Castel, A. D., McCabe, D. P., Roediger, H. L., III, & Heitman, J. L. (2007). The dark side of expertise: Domain specific memory errors. *Psychological Science*, 18, 3–5.

Chandrasekaran, B., Yi, H. G., Smayda, K. E., & Maddox, W. T. (2016). Effect of explicit dimensional instruction on speech category learning. *Attention, Perception, & Psychophysics*, *78*(2), 566-582.

Clapper, J. P. (2008). Category learning as schema induction. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A Festschrift for Gordon H. Bower*. New Jersey: Lawrence Erlbaum Associates.

Davis, T., Love, B. C., & Preston, A. R. (2012). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, *22*(2), 260-273.

Davis, T., Xue, G., Love, B. C., Preston, A. R., & Poldrack, R. A. (2014). Global neural pattern similarity as a common basis for categorization and recognition memory. *Journal of Neuroscience*, *34*(22), 7472-7484.

De Brigard, F. (2012). The Role of Attention in Conscious Recollection. *Frontiers in Psychology 3*. https://doi.org/10.3389/fpsyg.2012.00029.

De Brigard, F., Brady, T. F., Ruzic, L., & Schacter, D. L. (2017). Tracking the emergence of memories: A category-learning paradigm to explore schema-driven recognition. *Memory & Cognition*, *45*, 105-120.

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. Psychological Methods, 3(2), 186-205.

Dickerson, K. C., & Adcock, R. A. (2018). Motivation and memory. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, *1*, 1-36.

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, *17*(2), 224-230.

Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly journal of experimental psychology*, *68*(6), 1203-1222.

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 59–109). : Academic Press.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104-114.

Heit, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 712.

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. Topics in Cognitive Science, 1(1), 189-202.

Hsu, A. S., & Griffiths, T. E. (2010). Effects of generative and discriminative learning on use of category variability. In *32nd annual conference of the cognitive science society*.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological review*, *98*(3), 352.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment?. *Journal of experimental psychology: General*, *129*(2), 220.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299-312.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5), 1083.Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review, 25(1), 178-206.

Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 1211-1222.

Leotti, L. A., & Delgado, M. R. (2011). The inherent reward of choice. *Psychological science*, *22*(10), 1310-1318.

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & cognition*, *43*(2), 266-282.

Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences*, *108*(1), 55-60.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulleting and Review,* 9(4), 829-835.

Love, B. (2013). Categorization. In K.N. Ochsner and S.M. Kosslyn (Eds.) *Oxford Handbook of Cognitive Neuroscience*, 342-358. : Oxford University Press.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. Psychological review, 111(2), 309.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. Nature communications, 11(1), 1-11.

Makowski, D., Ben-Shachar, M. S., Chen, S. H., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767.

Mather, M., & Schoeke, A. (2011). Positive outcomes enhance incidental learning for both younger and older adults. Frontiers in neuroscience, 5, 129.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research*, *99*, 111-123.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of experimental psychology: learning, memory, and cognition*, *20*(4), 904.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *Journal of Neuroscience*, *35*(16), 6255-6264.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*(1), 39.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological review*, *101*(1), 53.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 21, 548–568.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.

Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1023.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?. *Journal of Memory and Language*, *60*(4), 437-447.

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803– 814.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology,* 31, 81–109.

Ruge, H., & Wolfensteller, U. (2010). Rapid formation of pragmatic rule representations in the human brain during instruction-based learning. *Cerebral Cortex*, *20*(7), 1656-1667.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, 133, 534-553.

Sakamoto, Y., & Love, B. C. (2010). Learning and retention through predictive inference and classification. *Journal of Experimental Psychology: Applied*, 16, 361-377.

Sakamoto, Y. (2012). Schematic influences on category learning and recognition memory. N. M. Seel (Ed.). *Encyclopedia of the Sciences of Learning*. Springer.

Seabrooke, T., Hollins, T. J., Kent, C., Wills, A. J., & Mitchell, C. J. (2019). Learning from failure: Errorful generation improves memory for items, not associations. *Journal of Memory and Language*, *104*, 70-82.

Smith, E. E., & Medin, D. L. (2013). *Categories and concepts*. Harvard University Press.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*(2-3), 167-196.

Tompary, A., & Thompson-Schill, S. L. (2021). Semantic influences on episodic memory distortions. *: General*.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, *114*(1), 152.

Yin, S., O'Neill, K., Brady, T.F., & De Brigard, F. (2019). The effect of category learning on recognition memory: A signal detection theory analysis. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society.* pp. 3165-3171.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002-3008.