Generalization Following Incidental and Intentional Category Learning

Stefania R. Ashby, Caitlin R. Bowman, Dagmar Zeithamova

University of Oregon

Author Note

Stefania R. Ashby, Department of Psychology, University of Oregon; Caitlin R. Bowman, Department of Psychology, University of Oregon; Dagmar Zeithamova, Department of Psychology, University of Oregon.

This work was supported by the National Institute on Aging Grant F32-AG-054204 awarded to Caitlin R. Bowman.

Correspondence concerning this article should be addressed to Dagmar Zeithamova, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403.

Email: dasa@uoregon.edu

Abstract

Memory generalization, the ability to link information across experiences to derive new knowledge, is a fundamental function of memory. Generalization is typically studied using tasks that probe the ability to extract commonalities across items, such as learning new concepts from a set of category examples. Less is known about the extent to which individuals spontaneously extract commonalities across items necessary for generalization when task goals emphasize item-specific information. In the present study, we used blends of faces to create face families and control for within- and between-category similarity. Memory generalization was tested in two experiments with contrasting goals at encoding. In the incidental category-learning experiment, participants underwent paired-associates training of face-name pairs. Faces had a unique first name, but category information was present as shared surnames. In the intentional category-learning experiment, participants underwent traditional feedback-based training of face families. Category bias in perceived similarity served as an implicit measure of category-learning while generalization of surnames to new faces served as an explicit measure. We found evidence of memory generalization irrespective of goals at encoding and prior to any explicit demands to make category judgments. Further, after incidental learning, where item-specific information was tested in addition to generalization, we found no evidence that spontaneous learning of commonalities across faces came at the expense of memory for face-specific information. Together, these results indicate that generalizable knowledge may form spontaneously alongside memory for specific details, enabling us to make distinct types of judgments from the same experiences.

Keywords: category learning, incidental, intentional, learning goals, perceived similarity, generalization

Generalization Following Incidental and Intentional Category Learning

The ability to link information across experiences to derive generalizable knowledge is a fundamental function of memory that is important for guiding behavior in novel situations. In the laboratory, memory generalization has traditionally been studied using tasks that probe the ability to extract commonalities across experiences, such as category learning (Posner & Keele, 1968). For example, participants may be shown a series of images and guess which category they belong to, learning the correct assignment of images into categories over time via corrective feedback. Generalization is then measured by testing the ability to apply category labels to new images. A complementary memory function—the ability to remember individual instances and specific details of past experience—has been studied in a distinct line of research using episodic memory tasks such as item recognition (Rugg & Yonelinas, 2003; Shepard, 1967) or paired-associate learning (Thorndike, 1908; Winocur & Weiskrantz, 1976). For example, participants may be shown arbitrary image-word pairs at encoding and then tested on their ability to retrieve the associated word for a given image. Utilizing distinct tasks focusing on either extracting commonalities or remembering specific details has been fruitful in teasing apart the cognitive and neural mechanisms implicated in memory generalization versus specificity (Bozoki, Grossman, & Smith, 2006; Knowlton & Squire, 1993; Nosofsky & Zaki, 1998; Poldrack et al., 2001; Reber, Stark, & Squire, 1998). However, daily life is rarely separated into situations for learning specific information versus learning commonalities in service of generalization. Rather, both forms of learning may proceed based on the same experiences. For example, when attending a wedding one may meet many new individuals and encode their names, but also notice commonalities across individuals and make inferences about who is related to whom.

Nevertheless, the extent to which individuals may acquire generalizable knowledge spontaneously is not well understood.

One approach that may be useful in answering this question is incidental category learning, where category information is present in some form but not emphasized during presentation of category examples (Aizenstein et al., 2000; Bozoki et al., 2006; Gabay, Dick, Zevin, & Holt, 2015; Kéri, Kálmán, Kelemen, Benedek, & Janka, 2001; Love, 2002; Reber, Gitelman, Parrish, & Mesulam, 2003). Participants are given a cover task to complete at encoding while viewing stimuli that contain an underlying category structure. Learning of category information is then evaluated by examining classification accuracy for never-before-seen category examples and accurate performance is taken as evidence of incidental generalization. Studies that tested both recognition and categorization for the same set of stimuli (often in distinct groups of participants) show above chance performance in both tasks (Bozoki et al., 2006; Kéri et al., 2001), suggesting that people may remember both specific instances and generalize based on the same experiences.

While prior work indicates that commonalities across experiences can be spontaneously extracted under learning conditions that do not explicitly require it, such generalization has been shown in a relatively narrow set of circumstances. First, prior incidental category learning studies have largely used a single category learning paradigm, also referred to as “A/not-A” (Ashby & Maddox, 2005; Ashby & O’Brien, 2005). In this paradigm, participants are exposed to examples of a single category during a study phase, and then informed that all of the stimuli they saw were members of a single category. During a test phase, participants are asked to decide whether a test stimulus is or is not a member of the same category. Category examples are mixed with non-categorical examples that typically do not form a coherent category in themselves.

Given that mechanisms of single category learning (A/not-A) differ from those involved in learning of contrasting categories (A/B) (Casale & Ashby, 2008; Zeithamova, Maddox, & Schnyer, 2008), it is unclear whether learning of contrasting categories can also proceed spontaneously. Furthermore, greater physical similarity among category examples but reduced physical similarity among non-examples may guide categorization at test (Zaki & Nosofsky, 2001), in extreme cases leading to apparent learning during test (Palmeri & Flanery, 1999). How spontaneous category learning proceeds when physical similarity does not clearly determine category membership remains unknown.

Another open question concerns to what degree people spontaneously link related information when directed to focus on differentiating details instead. Most commonly, instructions for encoding in incidental categorization tasks do not explicitly point participants' attention to stimulus commonalities or differences, such as when participants are asked to observe stimuli and identify their center (Kéri et al., 2001; Reber et al., 2003) or think about the appearance of the presented stimuli (Bozoki et al., 2006; Love, 2002). In one study (Aizenstein et al., 2000), participants' task was to respond as quickly as possible when a black-and-white stimulus changed to color, with categorical stimuli sharing the same color. This allowed the researchers to use a speeded response as an implicit measure of category generalization but also may point participants' attention to the commonalities across same-colored patterns rather than differences among stimuli. Past work has shown that top-down learning goals at encoding affect memory performance (Craig et al., 2016; Geiselman, Bjork, & Fishman, 1983; Murayama & Elliot, 2011). Furthermore, individuals perform better when the style of memory assessment is known prior to encoding and performance suffers when a mismatch occurs between test expectations and the actual test assessment (Hall, Grossman, & Elwood, 1976; Middlebrooks, Murayama, & Castel, 2017; Neely & Balota, 1981). To what degree spontaneous generalization occurs under conditions when participants are explicitly guided to differentiate individual items is unknown.

Open questions also remain regarding *when* memory generalization occurs and what type of memory representation it relies on. First, generalization may not be measurable until it is explicitly probed, either because it relies on specific memory traces with generalization judgments computed on-the-fly as needed, as exemplar models of categorization assume (Hintzman, 1984; Kruschke, 1992; Nosofsky, 1988), or because generalized representations are formed only in response to generalization demands (Carpenter & Schacter, 2017, 2018; Squire, 1992; Teyler & DiScenna, 1986; Winocur, Moscovitch, & Sekeres, 2007). Alternatively, generalized representations linking related experience may form during encoding, either instead of specific representations (Knowlton & Squire, 1993; Shohamy & Wagner, 2008; Zeithamova, Schlichting, & Preston, 2012) or alongside specific ones in distinct brain loci (Brunec et al., 2018; Collin, Milivojevic, & Doeller, 2015; Poppenk, Evensmoen, Moscovitch, & Nadel, 2013; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Schlichting, Mumford, & Preston, 2015; Schlichting, Zeithamova, & Preston, 2014). If so, generalization may manifest in behavior even before it is explicitly probed. These different views also make distinct predictions regarding the relationship between memory for individual experiences and generalization success. They may be positively related when relying on the same representations, negatively related if the formation of generalized representations comes at the expense of memory specificity or relatively unrelated if specific and generalized representations form in parallel in distinct loci of the brain.

In the present study, we tested the ability to generalize category knowledge obtained spontaneously (Experiment 1) or under explicit instruction (Experiment 2) when physical similarity alone could not be used to guide categorization judgments. Participants were shown faces that belonged to three categories (families), designated by a family surname. Face stimuli were created as blends of never-seen “parent” faces, resulting in increased physical similarity between faces that shared a parent (Figure 1). Physically similar faces that shared a so-called category-relevant parent were members of the same family while physically similar faces that shared a so-called category-irrelevant parent were members of different families, allowing us to dissociate the effect of category membership from physical similarity. In Experiment 1, participants were instructed to learn a full name for each face (first names unique for each face), thus emphasizing memory specificity and differences between all faces. This allowed us to determine the extent to which people can remember specific information *and* generalize based on the same experiences when learning goals at encoding emphasize memory specificity. In Experiment 2, the same faces were encoded under standard category learning instructions, emphasizing similarities among faces belonging to the same category. Generalization of category labels (family names) to new face-blend stimuli was assessed in both experiments to evaluate how learning of generalizable information proceeds with and without intention. To evaluate whether extraction of category knowledge may be detectable even before an explicit demand to generalize, we used an implicit generalization test by measuring perceived similarity between pairs of faces before and after the face encoding task. Prior research has shown that category knowledge biases perception, such that items within categories are perceived as more similar than items in different categories (Beale & Keil, 1995; Goldstone, 1994a; Goldstone, Lippa, & Shiffrin, 2001; Livingston, Andrews, & Harnad, 1998; Rosch & Mervis, 1975). Observing such a category bias in perceived similarity ratings after encoding (but before an explicit generalization test) would indicate that category representations have already been formed. Furthermore, if this category bias in perception reflects memory generalization, we would expect this measure to be positively related to an explicit measure of memory generalization. Finally, in Experiment 1 where measures of both specificity and generalization were obtained, we tested their correlation to determine whether they are based on the same representations, competing representations or co-existing representations.

Experiment 1: Incidental Category Learning

Methods

Participants. Forty-three healthy participants were recruited from the University of Oregon community via the university SONA research system and received course credit for their participation. All participants provided written informed consent, and all experimental procedures were approved by Research Compliance Services at the University of Oregon. Three participants were excluded from analyses for failing to make responses on more than 25% of categorization trials, and 1 additional participant was excluded due to incomplete data. As such, all analyses described below were carried out with 39 participants (M_{age} = 19.30, SD_{age} = 1.13, age range: 18-23 years, 21 females).

Stimuli. Stimuli were images of grayscale blended faces constructed by morphing two unaltered face images together using FantaMorph Version 5 by Abrosoft. It was important to our experimental design that the morphed faces look realistic and that pairs of face-blends could be equated for physical similarity between all blends, as differences in perceptual similarity after learning would be used to index memory generalization. Prior work has shown that category effects differ based on whether morphed faces are constructed from parents within one race

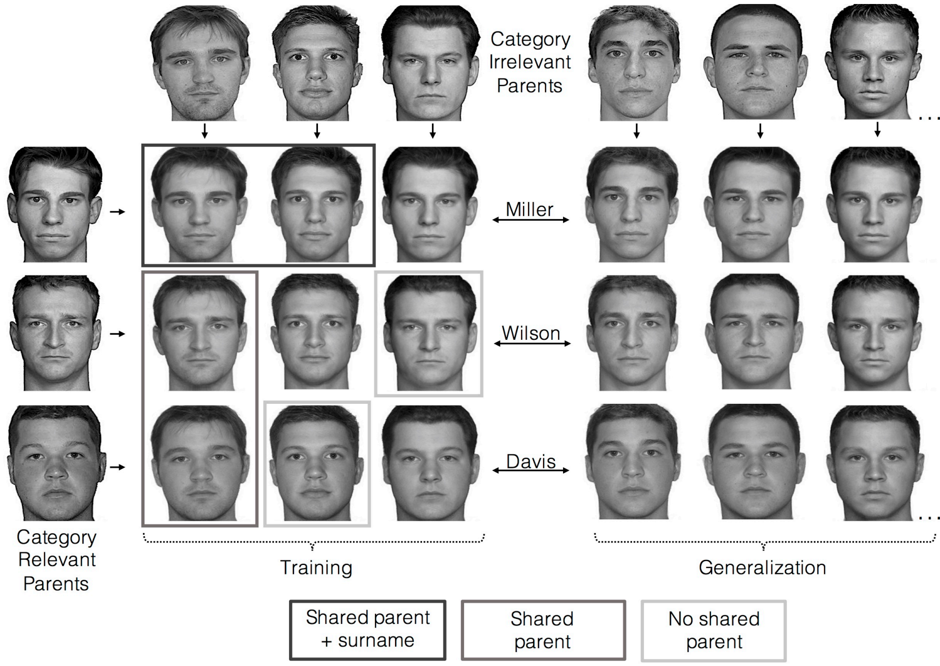

Figure 1. Example face-blend stimuli. Parent faces on the leftmost side are designated "category relevant parents" as these parents determined family membership—Miller, Wilson, or Davis—during training and generalization. Parent faces across the top are designated "category irrelevant parents" as these parents introduced physical similarity across families but did not determine categories. Three irrelevant parents were used for training. The rightmost three irrelevant parents are a subset of new faces used for generalization. Parent faces were never viewed by participants, only the resulting blended faces. The face blending procedure produced pairs of faces that shared a category-relevant parent and belonged to the same family (shared parent + surname; example indicated with dark grey box), pairs of faces that shared a category-irrelevant parent and belonged to different families (shared parent; example indicated with medium grey box). Non-adjacent pairs did not share a parent or a surname (example indicated with light grey boxes).

versus across two races (Levin & Angelone, 2002). Thus, we restricted all parent faces to be Caucasian to ensure that the resulting face-blend stimuli were comparably similar to all other faces with a shared parent. Finally, all parent faces were male to ensure that face-blends maintained a realistic appearance.

An example set of stimuli, including the parent faces, is presented in Figure 1. For each participant, three category-relevant parent faces and three category-irrelevant parent faces were randomly selected from a total set of twenty faces. Each of the three category-relevant parent faces was then individually morphed with each of the three category-irrelevant parent faces with equal weight given to each parent face (50/50 blend). The blending process produced facial stimuli that shared physical traits both within and between category boundaries such that physical similarity alone was not diagnostic of category membership. The resulting nine blended faces were then used as training stimuli during the face-name paired associates task described below. Faces blended with the same category-relevant parent shared perceptual similarities and belonged to the same family category while faces blended with the same category-irrelevant parent shared perceptual similarities but belonged to different families. The three category-relevant parent faces were also blended with the remaining fourteen faces from the larger face stimulus pool and used as generalization stimuli in a subsequent categorization task.

Procedure & statistical analysis. The experimental procedure consisted of six phases: passive initial exposure, pre-learning similarity ratings, paired-associates training (incidental category learning), post-learning similarity ratings, cued recall of face-name associations, and a surprise categorization task (Figure 2). Participants also completed two blocks of passive viewing of the face blends immediately before the pre- and post-learning similarity rating phases. These phases were included as a pilot of a future neuroimaging experiment, with no responses collected.

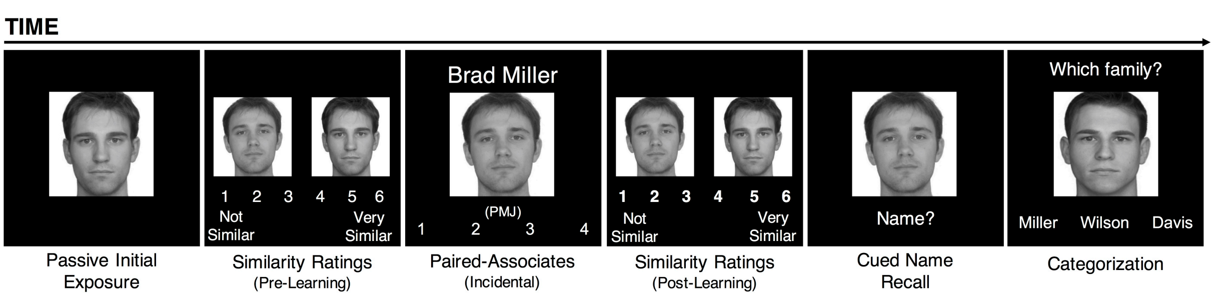

Figure 2. Incidental category learning task. Participants passively viewed all nine training stimuli once prior to completing any experimental tasks. Before and immediately after learning, similarity ratings were collected for all pair-wise comparisons of the training stimuli. To present category information incidentally, a face-name paired-associates task was used where face-blends were paired with full names. Each face had a unique first name but shared one of three family surnames. Prospective memory judgments (PMJ) were collected on each trial to facilitate participant engagement. A self-paced cued name recall test for face-name associations was completed prior to a surprise categorization task where the old training face-blends along with the new face-blends were categorized into families without feedback.

Passive initial exposure. To familiarize participants with the stimuli and give them an idea of the degree of similarity between all faces before collecting perceived similarity ratings during the first perceptual similarity task, participants first viewed each of the nine training stimuli individually (3s) once in a random order without any labels and without making any responses.

Pre-learning similarity ratings. To validate that participants were sensitive to the similarity structure among faces introduced by the blending process and to obtain baseline similarity ratings, participants rated the subjective similarity of pairs of faces to be used during the incidental category-learning phase. All possible 36 pairwise comparisons of the 9 training faces were presented (5s) and participants rated the similarity of the two faces on a scale from one to six: one indicating the two faces appeared very dissimilar and six indicating the two faces appeared very similar. Trials were then binned into three conditions, depending on whether pairs 1) shared a parent and a surname, 2) shared a parent face but did not share a surname, or 3) did not share a parent face. A one-way repeated measures ANOVA tested the effect of face pair in order to test whether faces that share a parent are perceived as more similar than those that do not share a parent. For all ANOVAs, a Greenhouse-Geisser correction for degrees of freedom (denoted as *GG* in all reported statistics) was used wherever Mauchly's test indicated a violation of the assumption of sphericity.

Paired-associates training (incidental category learning). To promote the acquisition of item-specific knowledge of individual face identities, participants completed paired associates training where each face-blend stimulus was paired with a full name (12 exposures across 3 training blocks). First names were unique to each face, while family surnames Miller, Wilson, or Davis were shared across faces that were blends of one of the three category-relevant parent faces. At each trial, participants studied the face-name pairs (4s) and then made a prospective memory judgement rating their confidence on being able to recall the face-name associations on a scale from one to four: one indicating they would definitely not remember the face-name association and four indicating they would definitely remember the face-name association. Prospective memory judgments were included to facilitate participant engagement with the observational learning task and were not considered further. While participants' goal at learning was to differentiate individual faces in order to learn face-specific information, the inclusion of the family surname provided an opportunity for incidental category learning if participants extracted commonalities across faces with a shared surname. The fact that surnames were repeated across faces or that there was a category structure among faces was not explicitly mentioned to participants.

Post-training similarity ratings. Perceived similarity ratings for all pairs of faces were also collected post-encoding, before explicit memory tests. These were used to measure a category bias on perceived similarity to probe spontaneous category learning. We reasoned that if participants spontaneously picked up on commonalities across faces with shared surnames and linked them into family representations, faces within a category may be perceived as more similar than those from different categories, even when physical similarity is equated across all faces. To test for a category bias on perception, we compared post-training similarity ratings for faces learned to belong to the same family (shared parent and surname) versus similar faces learned to belong to different families (shared parent but not surname) using a paired t-test. To test all effects of face-name training on perceived similarity ratings, we also conducted a 2 x 3 (timepoint [pre-training, post-training] x condition [shared parent and surname, shared parent, no shared]) repeated-measures ANOVA.

Cued recall of face-name associations. A self-paced cued recall test probed how well participants remembered face-specific information. Each training face was shown individually, and participants were asked to write down that individual's full name. Participants advanced the trials at their own pace but were not able to skip faces or go back and look at faces already named. Participants were encouraged to make their best guess as to the first name and surname of faces even if they were not confident in their memory. Using one-sample t-tests, memory for first names and surnames during cued recall was tested against chance (zero in a recall task).

Categorization. To explicitly test whether participants learned category information and whether that knowledge could be generalized to new faces, participants categorized the nine training stimuli as well as 42 never-seen face-blends, 14 from each family (i.e., 14 new blends of each of the three category-relevant parent faces). Face stimuli were viewed individually (4s), and participants were asked to select the appropriate family surname. The three family categories were presented on the screen for selection and participants indicated the category by pressing a button on the keyboard. No corrective feedback was provided and performance during categorization was indexed as the proportion correct, separately for the old training faces and for the new faces. One-sample t-tests were used to evaluate whether performance was significantly above what would be expected by chance (.33 for three categories). A paired-samples t-test was used to compare categorization success for old versus new items. Generalization scores (accuracy for categorizing the new faces) were correlated using Pearson's correlation coefficient with the category bias on perceived similarity, to test how strongly the implicit category bias measure predicts the explicit measure of generalization. We also correlated individual differences in memory for face-specific information (first name cued recall) with individual differences in generalization success to test whether spontaneous extraction of commonalities across related faces comes at the expense of memory for face-specific details.

Results

Similarity ratings. Mean similarity ratings for each condition are presented separately for pre-training (Figure 3A) and post-training (Figure 3B). Pre-training ratings demonstrated that participants were sensitive to the physical similarity differences among faces that we introduced with our face-blending procedure. A one-way, repeated measures ANOVA showed a significant effect of pair type ($F(1.46, 55.47) = 72.22$, $p < .001$, $\eta_p^2 = .655$, *GG*), driven by lower perceived similarity of faces that did not share a parent compared to faces that shared a parent (with and without shared surname, both $t > 10.65$, $p < .001$). For faces that shared a parent, ratings did not significantly differ when face pairs had the same or different (not yet presented) surname ($t(38) = 1.82$, $p = .077$).

To test the effect of training, we conducted a 2 x 3 (timepoint [pre-training, post-training] x condition [shared parent and surname, shared parent, no shared]) repeated-measures ANOVA.

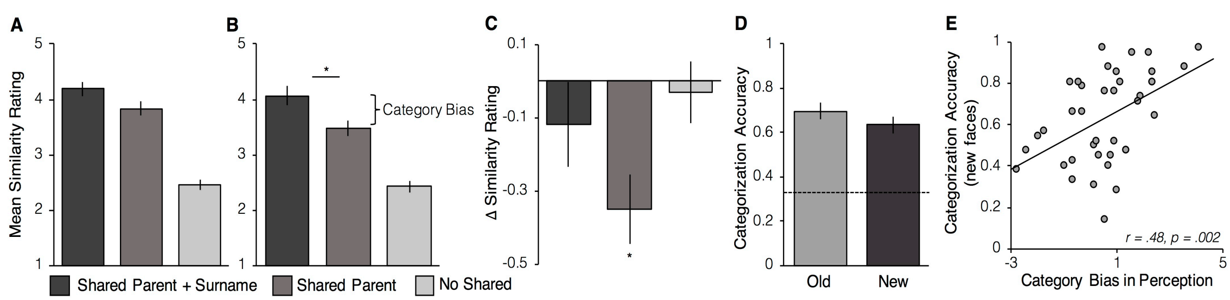

Figure 3. Incidental category learning results. ***A***. Average similarity ratings for faces that share a parent and surname, faces that only share a parent, and faces that don't share any parents before learning. ***B***. Average similarity ratings for the same pairwise comparisons after paired-associates learning. Asterisk represents a significant ($p < .05$) difference in post-learning similarity ratings for faces that belong to the same family vs. faces that share physical similarity but belong to different families (i.e. a category bias in perception). ***C***. Changes in similarity ratings from pre to post paired-associates learning. Asterisk denotes a significant ($p < .05$) decrease in perceived similarity for faces that share a parent but belong to different families. ***D***. Proportion of correct categorization responses for the training faces (old) as well as the novel face-blends (new). Dashed line represents chance (.33 for three categories). ***E***. Positive relationship between implicit (category bias in perception) and explicit (categorization accuracy for new faces) measures of memory generalization.

We found a significant main effect of timepoint ($F(1, 38) = 5.20$, $p = .028$, $\eta_p^2 = .120$) where similarity ratings for all faces were lower after training ($M_{pre} = 3.49$, $SD_{pre} = 0.51$; $M_{post} = 3.33$, $SD_{post} = 0.59$; $t(38) = 2.28$, $p = .028$, $d = 0.37$). We also found a significant main effect of condition ($F(1.28, 48.60) = 60.42$, $p < .001$, $\eta_p^2 = .614$, *GG*). Follow-up pairwise comparisons revealed that faces sharing any parent—shared parent and surname ($M = 4.13$, $SD = 0.92$) or shared parent ($M = 3.66$, $SD = 0.78$)—were rated significantly more similar to one another than those not sharing any parent ($M = 2.45$, $SD = 0.53$, both $t > 10.50$, $p < .001$, $d > 1.68$). Additionally, faces sharing a parent and surname were rated as more similar to one another than faces only sharing a parent ($t(38) = 2.23$, $p = .03$, $d = 0.36$). Critically, there was a significant interaction between timepoint and condition ($F(1.67, 63.37) = 4.21$, $p = .03$, $\eta_p^2 = .10$, *GG*). Follow-up pre-post comparisons within each condition (Figure 3C) revealed that this interaction was driven by a significant decrease in similarity ratings for faces sharing a parent but not a surname ($t(38) = -3.71$, $p = .001$, $d = -0.59$), but no significant decreases in similarity ratings in other conditions (both $t < -1.04$, $p > .30$, $d < -0.18$). Thus, changes in perceived similarity were affected by category membership.

The post-training category bias on perceived similarity, or the differences in perceived similarity of faces that shared a parent and surname versus those who shared a parent but not a surname, was used as an implicit measure of categorization. Across the group, we found that the category bias on perceived similarity was significantly greater than zero ($M_{difference} = 0.58$, $SD = 1.52$, $t(38) = 2.39$, $p = .022$). Although the difference in similarity ratings for faces with a shared parent and surname versus faces with only a shared parent was not statistically significant before training ($M_{difference} = .35$, $t(38) = 1.82$, $p = .077$), there was a numerical tendency. Parent faces were randomly selected for each participant and randomly assigned to serve as category-relevant

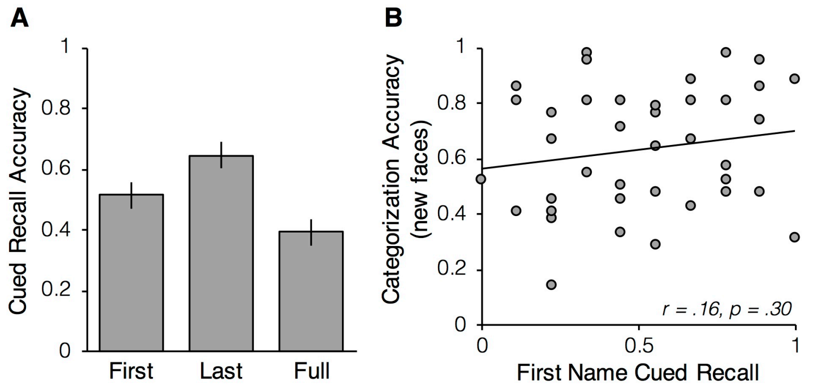

Figure 4. Specificity results. ***A***. Proportion of correct responses during the cued-recall test for first, last, and full names separately. Accuracy significantly (all p's < .01) above chance for all. ***B***. No significant relationship between first name cued recall and categorization accuracy for new faces.

or category-irrelevant parents for each subject. However, it seems that some of the category-relevant parent faces happened to be somewhat more salient, leading to a numerically greater pre-training similarity rating. Thus, we tested whether the post-training category bias on perceived similarity was reliably greater than pre-training. A 2 x 2 (timepoint [pre-training, post-training] x condition [shared parent and surname, shared parent]) repeated-measures ANOVA showed only a marginal interaction between timepoint and condition ($F(1, 38) = 2.87$, $p = .098$, $\eta_p^2 = .07$). We thus controlled for pre-training similarity rating differences in subsequent analyses.

Categorization. To directly test generalization of category knowledge, we examined accuracy in determining family membership during categorization (Figure 3D). One-sample t-tests revealed that participants were able to categorize the old training faces ($t(38) = 9.78$, $p < .001$, $d = 1.57$) and generalize category information to new faces ($t(38) = 8.66$, $p < .001$, $d = 1.39$) well above chance (.33 for three categories). A paired-samples t-test showed higher categorization accuracy for the old training faces ($M = .70$, $SD = .23$) than for the new ($M = .64$, $SD = .22$, $t(38) = 2.12$, $p = .04$, $d = 0.34$).

We next tested whether our implicit measure of generalization (the category bias on perceived similarity) was related to subsequent explicit generalization. Using a Pearson's correlation we found a significant positive relationship between the category bias on perceptual ratings and generalization accuracy ($r(37) = .48$, $p = .002$) such that the greater the category bias the better participants were able to generalize family surnames to new faces during categorization (Figure 3E). The category bias on perceived similarity in the post-training phase remained a significant predictor of subsequent categorization performance even when pre-training similarity ratings were considered (multiple regression: pre-training category bias $\beta = -.22$, $t(38) = -0.86$, $p = .40$; post-training category bias $\beta = .66$, $t(38) = 2.57$, $p = .01$), indicating that the post-training category bias predicting subsequent generalization was primarily due to category learning during training rather than pre-existing biases driven by random differences in parent face saliences.

Cued name recall. To test whether participants successfully learned face-name associations presented during training, we computed one-sample t-tests comparing first name, surname, and full name cued recall to zero—chance performance in a recall task. Results indicated that participants followed task instructions to learn individual names (first, last, full name recall, all $t > 6.5$, $p < .001$, Figure 4A).

To explore the relationship between memory for face-specific information and the ability to extract commonalities across faces in service of generalization, we examined correlations between first name recall and memory generalization performance during categorization. We did not find a significant correlation between first name recall and categorization accuracy for new faces ($r(37) = .17$, $p = .30$; Figure 4B). As would be expected, we found a significant positive correlation between better recall for surnames of the training faces and being able to assign these surnames to new faces ($r(37) = .69$, $p < .001$). A multiple regression model with first name and surname accuracies as predictors explained 48% of the total variance in categorization performance for new faces ($R^2 = .48$, $F(2, 36) = 16.63$, $p < .001$), with only the surname recall being a significant predictor of categorization success ($\beta = .73$, $t(38) = 5.59$, $p < .001$) while first name recall was not ($\beta = -.11$, $t(38) = -.82$, $p = .42$). Thus, superior associative memory for face-specific information (first name) was not related the ability to extract commonalities across faces in service of generalization.

Discussion

We hypothesized that individuals would be able to extract both item-specific and generalized information from the same experiences, even when extracting commonalities across experiences is not emphasized by task goals. We found evidence that information supporting both memory specificity and generalization was extracted simultaneously from the same experiences during encoding, and prior to explicit probes to make generalization decisions. The category bias on perceptual ratings after learning and subsequent success in categorizing new faces into the correct family categories demonstrated that participants spontaneously extracted commonalities across faces with shared surnames, which allowed them to successfully generalize family names to new faces. Furthermore, extracting commonalities across related faces did not appear to come at the expense for memory for face-specific information. These results align with the notion that events are represented simultaneously at multiple levels of specificity (Brunec et al., 2018; Collin, Milivojevic, & Doeller, 2015; McNaughton & Morris, 1987; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Schlichting, Zeithamova, & Preston, 2014) and suggest that specific and generalization judgments in this task relied on distinct representations.

Experiment 2: Intentional Category Learning

Top-down goals at learning can alter memory (Craig et al., 2016; Geiselman et al., 1983; Murayama & Elliot, 2011). Thus, there may be differences between generalization following incidental compared to intentional category learning. To test how the pattern of results in Experiment 1 compares to a traditional category-learning task, Experiment 2 used the same stimuli and category structure as Experiment 1, but employed feedback-based intentional category training, explicitly emphasizing similarities among faces within families.

Methods

Participants. Thirty-nine healthy participants were recruited from the University of Oregon community via the university SONA research system and received course credit for their participation. All participants provided written informed consent, and all experimental procedures were approved by Research Compliance Services at the University of Oregon. Four participants were excluded from analyses due to chance performance (accuracy $\leq$ .33) in categorizing the training faces. As such, all analyses described below were carried out with 35 participants (M_{age} = 20.43, SD_{age} = 2.58, age range: 18 – 32 years, 21 females).

Stimuli. We used the same face-blend stimuli as Experiment 1 (see Figure 1).

Procedure & statistical analyses. The experimental procedure was similar to the experimental procedure followed in Experiment 1 (see Figure 2) but with paired-associates face-name training being replaced by a traditional feedback-based intentional category learning task. At training, each face-blend stimulus was presented on the screen along with the three family labels as response options (16 exposures across 2 training blocks). Participants were instructed to indicate via a button press which family the face belongs to and received corrective feedback after each trial. Training accuracy was examined for the first and second half of training to show learning over time using a paired-sample t-test. Initial exposure, pre- and post-training similarity ratings, and the final categorization test were the same as in Experiment 1.

Results

Similarity ratings. Mean similarity ratings for each condition are presented separately for the pre-training (Figure 5A) and post-training phases (Figure 5B). Pre-training similarity ratings differed significantly among pair types ($F(2, 68) = 58.74$, $p < .001$, $\eta_p^2 = .63$), driven by lower perceived similarity for faces that did not share a parent compared to those that shared a parent (with or without same surname, both $t > 9.17$, $p < .001$). Faces that shared a parent were perceived as equally similar to one another irrespective of whether they also shared the same (not yet presented) surname ($t(34) = -0.17$, $p = .87$). A 2 x 3 (timepoint [pre-training, post-training] x condition [shared parent and surname, shared parent, no shared]) repeated-measures ANOVA testing the effect of category training on perceived similarity ratings revealed no main effect of timepoint [$F(1, 34) = 0.04$, $p = .85$, $\eta_p^2 = .001$], a significant main effect of condition (F(1.63,

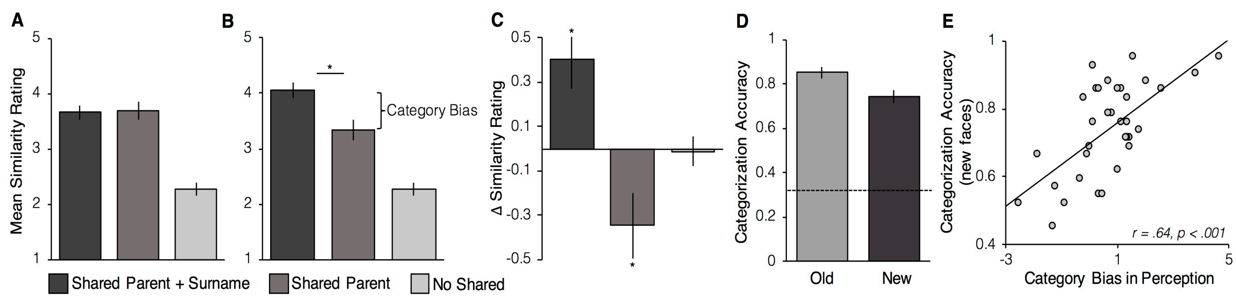

Figure 5. Intentional category learning results. ***A***. Average similarity ratings for faces that share a parent and surname, faces that only share a parent, and faces that don't share any parents before learning. ***B***. Average similarity ratings for the same pairwise comparisons after traditional feedback-based category learning. Asterisk represents a significant (p < .05) difference in post-learning ratings for faces within and between family categories (category bias in perception). ***C***. Changes in similarity ratings from pre to post intentional category learning. Asterisks denote a significant (p < .05) increase in perceived similarity for faces that share a parent and surname and a significant decrease in perceived similarity for faces that only share a parent. ***D***. Proportion of correct categorization responses for training faces (old) and novel face-blends (new). Dashed line represents chance (.33 for three categories). ***E***. Positive relationship between implicit (category bias in perception) and explicit (categorization accuracy for new faces) measures of memory generalization.

55.38) = 61.21, p < .001, η_p^2 = .64, *GG*), and a significant interaction between timepoint and condition (F(1.64, 55.88) = 11.85, p < .001, η_p^2 = .25, *GG*). Follow-up pre-post comparisons within each condition (Figure 5C) revealed that this interaction was driven by both a significant *increase* in similarity ratings for faces sharing a parent and a surname (t(34) = 3.02, p = .005, d = 0.51) and a significant *decrease* in similarity ratings for faces only sharing a parent and not a surname (t(34) = -2.33, p = .026, d = -0.39). There was no significant change in similarity ratings for faces that did not share a parent (t(34) = -.18, p = 0.86, d = -0.03). Testing category bias on perceived similarity, pairs of faces sharing a parent and surname were perceived as significantly more similar than faces that shared a parent but not a surname after training ($M_{difference}$ = .72, SD = 1.41, t(34) = 3.02, p = .005, d = 0.51) but not before training ($M_{difference}$ = -.03, SD = 0.97, t(34) = -0.17, p = .87, d = 0.02). The category bias differed significantly between pre- and post-training (F(1, 34) = 16.11, p < .001, η_p^2 = .32).

Intentional category learning. Given that participants were providing categorization response throughout training, we were able to track their learning during this phase. Participants were able to learn the three family categories with an average of 76% accuracy (t(34) = 17.66, p < .001, d = 3.01). Categorization accuracy improved across training, from 66% in the first half to 85% in the second half (t(34) = 9.72, p < .001, d = 1.63), demonstrating learning over time.

Categorization. To directly test generalization of category knowledge, we examined accuracy in determining family membership for both the old training face-blends as well as the new face-blends (Figure 5D). Participants' performance was well above chance for both the old training faces (t(34) = 18.56 , p < .001, d = 3.06) and for the new faces (t(34) = 18.12, p < .001, d = 3.15). A paired-samples t-test showed higher categorization accuracy for the old training faces (M = .85, SD = .17) than for the new (M = .74, SD = .13, t(34) = 5.48, p < .001 , d = 0.93).

We next tested whether our implicit measure of generalization (the category bias on perceived similarity ratings) was related to subsequent explicit generalization success. Using Pearson's correlation we found a significant positive relationship between the category bias on perceptual ratings and generalization accuracy ($r(33) = .64$, $p < .001$; Figure 5E). The category bias on perceived similarity in the post-training phase was a significant predictor of subsequent generalization performance even when pre-training similarity ratings were considered (multiple regression: pre-training differences in perceptual similarity $\beta = .30$, $t(34) = 1.80$, $p = .08$; post-training category bias $\beta = .46$, $t(34) = 2.75$, $p = .01$).

We found an advantage for categorizing old items versus generalizing to new items, suggesting that participants to some degree relied on memory for specific training instances in addition to generalized category representations. While this old advantage was found following both incidental and intentional training, we hypothesized that the old advantage would be greater after paired-associate learning (incidental category learning) than intentional learning, which would indicate that generalization after incidental category learning was more difficult than after intentional category learning. Although the two experiments were not designed as a single two-group experiment and procedures were not completely matched (participants had more exposure to the faces during the intentional than incidental training) we reasoned that their comparison might still be informative. We compared the effect of training version on categorization accuracy for old and new items by conducting a 2 x 2 (training version [incidental category learning, intentional category learning] x stimulus type [old, new]) mixed-effects ANOVA with categorization accuracy as the dependent variable. We found a significant main effect of training version ($F(1, 72) = 9.63$, $p = .003$, $\eta_p^2 = .12$) where participants who underwent intentional category training had higher overall categorization accuracy ($M = .80$, $SD = .14$) than participants who underwent incidental category training (M = .67, SD = .21). However, there was no significant interaction between training version and stimulus type ($F(1, 72) = 1.64$, $p = .21$, $\eta_p^2 = .02$). Thus, the advantage of categorizing old faces versus generalizing surnames to new faces did not vary as a function of training style.

General Discussion

Memory allows us to link across related experiences to build generalized knowledge. It has long been known that individuals can learn a broad array of category structures when explicitly instructed to do so (Ashby & Maddox, 2005, 2011; Hintzman, 1986; Kemler-Nelson, 1984; Knowlton & Squire, 1993; Nosofsky, 1986; Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968, 1970; Smith, Patalano, & Jonides, 1998). While memory generalization has typically been studied using tasks that explicitly focus on extracting commonalities across experiences, in the real world we must extract both commonalities and specific details from the same experiences. However, few studies have shown how individuals can extract category structure incidentally and those that have typically have done so under conditions where category structure can be determined by patterns of physical similarity (Aizenstein et al., 2000; Bozoki et al., 2006; Kéri et al., 2001; Love, 2002; Reber et al., 2003). Here we tested the prediction that individuals spontaneously form category representations even when category information is presented incidentally, when category membership is dissociable from physical similarity, and when task instructions emphasize individuating category members. Evidence for category learning was found irrespective of learning goals. First, individuals were able to generalize category labels to new examples even though physical similarity among stimuli did not unambiguously determine category membership. A category bias in perceived similarity judgments—which indirectly measured category learning prior to an explicit categorization test—indicated that category learning occurred incidentally during encoding rather than in response to generalization demands at test. The advantage for categorizing old items versus generalizing to new items was comparable between incidental and intentional learning, indicating that generalization after incidental learning was not disproportionately more difficult. Generalization did not appear to be based solely on memory for trained faces nor did it appear to come at the expense of memory for face-specific information, as we did not find a reliable correlation between face-specific information recall and generalization performance. Instead, learning of commonalities appeared to proceed spontaneously even when not explicitly instructed, and in parallel with learning of specific information.

While our results indicate that people can generalize irrespective of explicit goals at encoding, generalization performance was higher in Experiment 2 in which participants learned categories intentionally, consistent with prior findings that top-down goals affect memory performance (Craig et al., 2016; Geiselman et al., 1983; Murayama & Elliot, 2011). We also found distinct patterns of pre- to post-learning changes in perceived similarity in our two experiments. Intentional category learning was associated with increases in perceived similarity from pre- to post-learning for faces sharing a category label and decreases for equally similar faces across categories, consistent with prior studies (Beale & Keil, 1995; Goldstone, 1994a, 1994b; Goldstone et al., 2001; Rosch & Mervis, 1975). These shifts in perceived similarity may reflect allocation of selective attention to features that are category-relevant while diverting attention away from category-irrelevant features (Goldstone & Steyvers, 2001; Kruschke, 1996; Nosofsky, 1991). In contrast, incidental learning was associated with a decrease in overall similarity ratings across all types of comparisons from pre- to post-learning. This decrease could reflect learning-related differentiation of representations to minimize confusability and interference (Chanales, Oza, Favila, & Kuhl, 2017; Favila, Chanales, & Kuhl, 2016; Hulbert & Norman, 2015; Kim, Norman, & Turk-Browne, 2017; Lohnas et al., 2018). For example, neural patterns in the hippocampus for items that shared a common associate became more dissimilar compared to items that did not share a link, with the degree of differentiation predicting subsequent resistance to interference (Favila et al., 2016). Findings from our incidental task newly show that individuals still successfully encode similarity across related items despite the overall differentiation of representations. Together these results suggest that learning goals at encoding modulate the relative strength of different memory representations that are constructed simultaneously rather than leading to the formation of one type of memory representation at the expense of the other.

The inclusion of similarity ratings also allowed us to address the question of *when* category representations may emerge. Some theories assume that generalization decisions are computed on-the-fly from specific memories (Hintzman, 1984; Kruschke, 1992; Nosofsky, 1988). Another possibility is that people form generalized representations, but do so in response to generalization demands (Carpenter & Schacter, 2017, 2018; Squire, 1992; Teyler & DiScenna, 1986; Winocur et al., 2007). Instead, our finding of a category bias in perceived similarity *after* learning but *before* the explicit generalization test indicates that memory generalization in our task occurred spontaneously at encoding (*see also* Shohamy & Wagner, 2008; Zeithamova, Dominick, & Preston, 2012). Our results also extend prior studies on changes in perceived similarity as a result of explicit category learning (Goldstone, 1994b, 1994a; Livingston et al., 1998) to the domain of incidental category learning. The emergence of a category bias under incidental learning indicates that the mere presence of a shared piece of information can bias perceived similarity in many participants. Further, the positive relationship between individual differences in category bias in perceived similarity and subsequent generalization test indicates that both measures may reflect the same memory representations.

The notion that people may simultaneously remember specific experiences and extract commonalities across experiences has a long tradition. For example, multiple memory systems views posit that specificity and generalization judgments are based on distinct representations, formed by separate neural systems (Foerde, Knowlton, & Poldrack, 2006; Nomura & Reber, 2008; Poldrack & Packard, 2003; Smith & Grossman, 2008). Traditionally, these views have distinguished between the hippocampus, well suited to rapidly learn details of individual episodes (McClelland, McNaughton, & O'Reilly, 1995; Norman, 2010; O'Reilly & Norman, 2002; O'Reilly & Rudy, 2001), and striatum (Maddox & Filoteo, 2001; Myers et al., 2003; Seger & Cincotta, 2005) or cortex (Aron et al., 2004; McDonald & White, 1993; Seger & Miller, 2010) which incrementally extract commonalities across experiences to form generalized representations. More recent evidence indicates that the hippocampus itself is capable of rapidly forming generalized representations in both episodic memory tasks (Schlichting et al., 2015, 2014; Shohamy & Wagner, 2008) and in concept learning (Bowman & Zeithamova, 2018). Such generalized representations may form at the expense of memory for specific details, either because they are in competition at the time of encoding (Knowlton & Squire, 1993; Shohamy & Wagner, 2008; Zeithamova, Schlichting, et al., 2012) or because linking memories in response to generalization demands results in a loss of specificity (Carpenter & Schacter, 2017, 2018). Other work indicates that the hippocampus may form memory representations at multiple levels of specificity in parallel, perhaps represented across distinct hippocampal subfields (McNaughton & Morris, 1987; Schapiro et al., 2017; Schlichting et al., 2014) and/or along the long-axis of the hippocampus (Brunec et al., 2018; Collin et al., 2015; Poppenk et al., 2013; Schlichting et al., 2015). Our data are most consistent with the view that memory representations are formed at multiple levels of specificity without competition, although we cannot rule out the possibility that a specificity-generalization trade-off could emerge (Carpenter & Schacter, 2017, 2018) if memory for face-specific information was tested following generalization.

Understanding how individuals extract commonalities across experiences while also remembering unique aspects of each experience is key to understanding healthy memory function. The present study showed that individuals can spontaneously extract category information even when categories are not differentiated based on physical similarity alone, and when the task at hand emphasizes differentiating category members. Further, we showed that this spontaneous generalization need not come at the expense of memory specificity, as there was no trade off between memory for item-specific information and the ability to later generalize. We found that participants' category knowledge affected perceived similarity between items, and that this category bias in perception was present for both incidental and intentional learning of categories. Lastly, compared to intentional learning, categorization performance was worse overall following incidental learning, but the ability to generalize was not disproportionately affected. Together, these results indicate that generalizable knowledge may form spontaneously alongside memory for specific details, enabling us to make distinct types of judgments from the same experiences.

References

Aizenstein, H., MacDonald, A., Stenger, V., Nebes, R., Larson, J., Ursu, S., & Carter, C. (2000). Complementary category learning systems identified using fMRI. *Journal of Cognitive Neuroscience*, *12*(6), 977–987.

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology*, *92*, 1144–1152. https://doi.org/10.1152/jn.01209.2003

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*(2), 83–89. https://doi.org/10.1016/j.tics.2004.12.003

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

Bowman, C. R., & Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *The Journal of Neuroscience*, *38*(10), 2811–2817. https://doi.org/10.1523/JNEUROSCI.2811-17.2018

Bozoki, A., Grossman, M., & Smith, E. E. (2006). Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia*, *44*(5), 816–827. https://doi.org/10.1016/j.neuropsychologia.2005.08.001

Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z. X., … Moscovitch, M. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current Biology*, *28*(13), 2129–2135. https://doi.org/10.1016/j.cub.2018.05.016

Carpenter, A. C., & Schacter, D. L. (2017). Flexible retrieval: When true inferences produce false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 335–349.

Carpenter, A. C., & Schacter, D. L. (2018). False memories, false preferences: Flexible retrieval mechanisms supporting successful inference bias novel decisions. *Journal of Experimental Psychology: General*, *147*(7), 988–1004. https://doi.org/10.1037/xge0000391

Casale, M. B., & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Perceptual Psychophysiology*, *70*(6), 983–999. https://doi.org/10.1007/s11103-011-9767-z.Plastid

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Current Biology*, *27*, 1–11. https://doi.org/10.1016/j.cub.2017.06.057

Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*(11), 1562–1564. https://doi.org/10.1038/nn.4138

Craig, M., Butterworth, K., Nilsson, J., Hamilton, C. J., Gallagher, P., & Smulders, T. V. (2016). How does intentionality of encoding affect memory for episodic information? *Learning and Memory*, *23*(11), 648–659. https://doi.org/10.1101/lm.041491.115

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*, 11066. https://doi.org/10.1038/ncomms11066

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences*, *103*(31), 11778–11783. https://doi.org/10.1073/pnas.0602659103

Gabay, Y., Dick, F. K., Zevin, J. D., & Holt, L. L. (2015). Incidental auditory category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1124–1138. https://doi.org/10.1037/xhp0000073

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*(1), 58–72. https://doi.org/10.1037/0096-3445.112.1.58

Goldstone, R. L. (1994a). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200. https://doi.org/10.1037/0096-3445.123.2.178

Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125–157.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43. https://doi.org/10.1016/S0010-0277(00)00099-8

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139.

Hall, J. W., Grossman, L. R., & Elwood, K. D. (1976). Differences in encoding for free recall vs. recognition. *Memory & Cognition*, *4*(5), 507–513. https://doi.org/10.3758/BF03213211

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological*

Review, *93*(4), 411–428.

Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

Kemler-Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior*, *23*(6), 734–759.

Kéri, S., Kálmán, J., Kelemen, O., Benedek, G., & Janka, Z. (2001). Are Alzheimer's disease patients able to learn visual prototypes? *Neuropsychologia*, *39*(11), 1218–1223. https://doi.org/10.1016/S0028-3932(01)00046-X

Kim, G., Norman, K. A., & Turk-Browne, N. B. (2017). Neural differentiation of incorrectly predicted memories. *The Journal of Neuroscience*, *37*(8), 2022–2031. https://doi.org/10.1523/JNEUROSCI.3272-16.2017

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749. https://doi.org/10.1126/science.8259522

Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Kruschke, John K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*(2), 225–247. https://doi.org/10.1080/095400996116893

Levin, D. T., & Angelone, B. L. (2002). Categorical perception of race. *Perception*, *31*(5), 567–578. https://doi.org/10.1068/p3315

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning Memory and*

Cognition, *24*(3), 732–753. https://doi.org/10.1037/0278-7393.24.3.732

Lohnas, L. J., Thesen, T., Doyle, W. K., Devinsky, O., Duncan, K., & Davachi, L. (2018). Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proceedings of the National Academy of Sciences*, *115*(31), E7418–E7427. https://doi.org/10.1073/pnas.1717088115

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

Maddox, W. T., & Filoteo, J. V. (2001). Striatal contributions to category learning: Quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society*, *7*(6), 710–727. https://doi.org/10.1017/s1355617701766076

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, *107*(1), 3–22.

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences*, *10*(10), 408–415.

Middlebrooks, C. D., Murayama, K., & Castel, A. D. (2017). Test expectancy and memory for important information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1291–1296. https://doi.org/10.1097/CCM.0b013e31823da96d.Hydrogen

Murayama, K., & Elliot, A. J. (2011). Achievement motivation and memory: Achievement goals differentially influence immediate and delayed remember-know recognition memory. *Personality and Social Psychology Bulletin*, *37*(10), 1339–1348. https://doi.org/10.1177/0146167211410575

Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., … Schwartz, R. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience*, *15*(2), 185–193. https://doi.org/10.1162/089892903321208123

Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & Cognition*, *9*(3), 283–300. https://doi.org/10.3758/BF03196962

Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 279–291. https://doi.org/10.1016/j.neubiorev.2007.07.006

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: Revisiting the complementary learning systems model. *Hippocampus*, *20*(11), 1217–1227. https://doi.org/10.1002/hipo.20855.How

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27. https://doi.org/10.1037/0096-1523.17.1.3

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals. *Psychological Science*, *9*(4), 247–255.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*(12), 505–510. https://doi.org/10.1016/S1364-6613(02)02005-3

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345.

Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, *10*(6), 526–530.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251. https://doi.org/10.1016/S0028-3932(02)00157-4

Poldrack, R., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550. https://doi.org/10.1038/35107080

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*(5), 230–240. https://doi.org/10.1016/j.tics.2013.03.005

Posner, M. I., Goldsmith, R., & Welton, K. E. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, *73*(1), 28–38.

https://doi.org/10.1037/h0024135

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363.

Posner, M. I., & Keele, S. W. (1970). The retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2), 304–308. https://doi.org/10.1037/h0028558

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, *15*(4), 574–583. https://doi.org/10.1162/089892903321662958

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, *5*, 420–428. https://doi.org/10.1101/lm.5.6.420

Rosch, E., & Mervis, C. B. (1975). Family resemblances. *Cognitive Psychology*, *7*, 573–605. https://doi.org/10.1186/gb-2002-3-12-reports0063

Rugg, M. D., & Yonelinas, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *7*(7), 313–319. https://doi.org/10.1016/S1364-6613(03)00131-1

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 1–10. https://doi.org/10.1038/ncomms9151

Schlichting, M. L., Zeithamova, D., & Preston, A. R. (2014). CA1 subfield contributions to memory integration and inference. *Hippocampus*, *24*(10), 1248–1260. https://doi.org/10.1002/hipo.22310

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience*, *25*(11), 2941–2951. https://doi.org/10.1523/JNEUROSCI.3401-04.2005

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219. https://doi.org/10.1146/annurev.neuro.051508.135546

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

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*, 378–389. https://doi.org/10.1016/j.neuron.2008.09.023

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 249–264. https://doi.org/10.1016/j.neubiorev.2007.07.009

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*(2–3), 167–196. https://doi.org/10.1016/S0010-0277(97)00043-7

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231. https://doi.org/10.1037/0033-295X.99.3.582

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral*

Neuroscience, *100*(2), 147–154. https://doi.org/10.1037/0735-7044.100.2.147

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review*, *15*(2), 122–138. https://doi.org/10.1037/h0073570

Winocur, G., Moscovitch, M., & Sekeres, M. (2007). Memory consolidation or transformation: Context manipulation and hippocampal representations of memory. *Nature Neuroscience*, *10*(5), 555–557. https://doi.org/10.1038/nn1880

Winocur, G., & Weiskrantz, L. (1976). An investigation of paired-associate learning in amnesic patients. *Neuropsychologia*, *14*(1), 97–110. https://doi.org/10.1016/0028-3932(76)90011-7

Zaki, S. R., & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective and Behavioral Neuroscience*, *1*(4), 344–359. https://doi.org/10.3758/CABN.1.4.344

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179. https://doi.org/10.1016/j.neuron.2012.05.010

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: Evidence from brain imaging and behavior. *The Journal of Neuroscience*, *28*(49), 13194–13201. https://doi.org/10.1523/JNEUROSCI.2915-08.2008

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Frontiers in Human Neuroscience*, *6*, 1–14. https://doi.org/10.3389/fnhum.2012.00070