# Journal of Experimental Psychology: General

# **Semantic Influences on Episodic Memory Distortions**

Alexa Tompary and Sharon L. Thompson-Schill Online First Publication, January 21, 2021. http://dx.doi.org/10.1037/xge0001017

CITATION

Tompary, A., & Thompson-Schill, S. L. (2021, January 21). Semantic Influences on Episodic Memory Distortions. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0001017

https://doi.org/10.1037/xge0001017

# Semantic Influences on Episodic Memory Distortions

Alexa Tompary and Sharon L. Thompson-Schill Department of Psychology, University of Pennsylvania

Prior knowledge has long been known to shape new episodic memories. However, it is less clear how prior knowledge can scaffold the learning of a new class of information, and how this can bias memory for the episodes that contributed to its acquisition. We aimed to quantify distortions in episodic memories resulting from the use of prior category knowledge to facilitate learning new information. Across 4 experiments, participants encoded and retrieved image-location associations. Most members of a category (e.g., birds) were located near each other, such that participants could leverage their prior category knowledge to learn the spatial locations of categories as they encoded specific image locations. Critically, some typical and atypical category members were in random locations. We decomposed location memory into 2 measures: error, a measure of episodic specificity; and bias toward other category members, a measure of the influence of newly-learned information about category locations. First, we found that location memory was more accurate for images whose locations were spatially consistent with their category membership. Second, when images were spatially inconsistent (i.e., in random locations), retrieval of typical category members was more biased toward their category's location relative to atypical ones. These effects replicated across 3 experiments, disappeared when images were not arranged by category, and were stronger than effects observed with images arranged by visual similarity rather than category membership. Our observations provide compelling evidence that memory is a reconstruction of multiple sources of prior knowledge, new learning, and memory for specific events.

Keywords: episodic memory, semantic memory, memory reconstruction, distortion

Supplemental materials: https://doi.org/10.1037/xge0001017.supp

A fundamental way that humans acquire new information is by extracting relevant commonalities across memories of distinct events, or episodes (McClelland et al., 1995). However, much of the research in this area neglects the critical notion that new experiences are made up of recombinations of objects, places, and people for which we already have rich semantic knowledge. In other words, the episodes that contribute to new information are not formed in a vacuum, but rather are scaffolded by prior knowledge. As an example, imagine a new birding enthusiast who, after observing many different nests over the course of many nature walks, has come to understand that the color and speckling of eggs is often related to the species of bird—where similar species may produce similarly colored eggs. One day, she finds a bird's nest

Alexa Tompary D https://orcid.org/0000-0001-7735-3849

This research was funded by NIH F32 NS108511 to Alexa Tompary and NIH R01 DC009209 to Sharon L. Thompson-Schill.

A draft of this article and its online supplemental material was posted to PsyArXiv. Results from Experiments 1 and 2 were reported in the Proceedings of the 41st Annual Meeting of the Cognitive Science Society; Experiments 3 and 4 have never been reported. Data, analysis code, and stimuli are available at https://osf.io/8ugek.

Correspondence concerning this article should be addressed to Alexa Tompary, Department of Psychology, University of Pennsylvania, 3710 Hamilton Walk, Philadelphia, PA 19104, United States. Email: atompary@ sas.upenn.edu filled with unusually colored, bright blue eggs. Because bird eggs are usually not so vibrant, would the knowledge that the nest belongs to the more typical robin, compared with the more unusual emu, have consequences for how vibrantly the blue of the eggs is remembered? In other words, how might she integrate prior knowledge about birds, like whether the bird is a typical or atypical category member, with newly learned information about egg coloring, to shape a specific episodic memory? How the organization of our prior knowledge can influence new learning is critical to understand, because as new knowledge is acquired, it may shape how the episodes that contribute to this knowledge are remembered. Yet, in the context of new learning, the relationship between prior knowledge and new episodic memories is relatively understudied.

In the current experiments, we investigated how prior category knowledge influenced new episodic memories for the spatial locations of images, as participants incrementally learned that spatial locations sometimes were predicted by an image's category membership (i.e., many but not all images of birds were clustered together). Critically, these locations were not intrinsically related to the images (i.e., nothing about the concept "cardinal" implies that it should be located on the bottom right corner of a computer screen), but participants could learn the general location of a category by integrating over the locations of several category members. This category-location mapping is our operationalization of "new information" that can be learned using prior knowledge of category membership. We use this novel protocol to investigate how category knowledge—specifically, category typicality—can distort episodic memories in the context of learning this new information.

The experiments were motivated by three distinct lines of research: (a) spreading activation theory as an explanation for category typicality effects in episodic memory, (b) predictions of the category adjustment model (CAM) concerning how category typicality affects memory reconstruction, and (c) theories concerning how relational encoding shapes memory for distinctive or atypical items. As these literatures have contributed to our understanding of how prior knowledge modulates memories for individual episodes, we build on them to ask how category typicality can affect memories that are formed in the context of newly learned information. Below, we describe these three literatures, focusing on their connection with the predictions of the current experiments.

#### Typicality Influences Spreading Activation in Memory

Our prior semantic knowledge naturally comprises a rich, dimensional structure. There are many well-studied aspects of the structure of conceptual knowledge—like the relative frequency or familiarity of different concepts (e.g., McCloskey, 1980), or their hierarchical organization (e.g., Collins & Quillian, 1969)—that may influence how it affects memory. Here, we focus on the notion that categories have a graded internal structure, such that category members continuously range in their typicality, or whether they are good or less good examples of their category. Typical category members are defined as items that share the greatest number of features with other category members and the least number of features with members of other categories (Rosch et al., 1976).

Spreading-activation theory successfully explains many effects of this graded structure on behavior. This theory posits that concepts are nodes situated in a large network, where the links between nodes represent shared properties of concepts. Because typical category members share a greater number of features with other category members, leading to stronger links in the network, a typical category member is more likely to activate information about the rest of the category (Collins & Loftus, 1975). This theory accounts for ways in which category typicality shapes behavior across countless domains, including (but not limited to) categorization speed (Murphy & Brownell, 1985; Rips et al., 1973), production/ item dominance (Battig & Montague, 1969; Mervis et al., 1976), and feature induction (Dunsmoor & Murphy, 2014; Osherson et al., 1990; Rips, 1975). Of particular relevance to the current experiments, this theory explains various influences of category typicality in episodic memory. For instance, participants are more likely to cluster their recall of typical category members relative to atypical ones (Bousfield et al., 1958). On the other hand, when category members of high and low typicality are withheld from an encoded word list, the more typical category members are more likely to be falsely remembered (Roediger & McDermott, 1995; Smith et al., 2000).

Taken together, this work suggests that typical category members are more strongly associated with their category relative to atypical ones, and thus they may be more influenced by relevant category knowledge in the context of episodic memory. What is currently unknown is whether—in a situation where category knowledge can facilitate learning—typicality shapes how memory for new episodes is distorted. In the present experiments, the spreading-activation theory leads to the prediction that when learning novel image-location associations, memory for typical category members will be more influenced by the spatial locations of other category members.

# **Typicality Effects in Memory Reconstruction**

Episodic retrieval has been well characterized as an act of reconstruction, supported by the integration of different sources of information (Bartlett, 1932; Brewer & Nakamura, 1984; Schacter et al., 1998). Indeed, it is well known that episodic memories can be easily influenced by many sources of information, including retroactive interference, gist-based false memory driven by interference with the surrounding context, and more (E. F. Loftus & Palmer, 1974; Roediger & McDermott, 1995). The category adjustment model (Huttenlocher et al., 1991, 2000) formalizes this notion by positing that integrating imperfect estimates of an encoded event with categorical information maximizes average accuracy. This model leads to two predictions: First, if a new event is consistent with category information, then integrating the two sources of information will improve the "signal" of a memory, resulting in more accurate retrieval. Such integration provides an explanation for a large and diverse body of work showing that prior knowledge facilitates new memory formation and retrieval (Alba & Hasher, 1983; Bransford & Johnson, 1972; Brewer & Treyens, 1981; Schulman, 1974).

Second, CAM predicts systematic errors if there are discrepancies between a new memory and prior knowledge, and the direction and extent of these errors depend on the uncertainty of the encoded representation and of the relevant category information. Specifically, estimates of category members are often remembered as more typical to their category than they actually are-known as the central tendency bias or schema effect (Bartlett, 1932; Hollingworth, 1910; Poulton, 1979). A wide range of studies provides evidence for this model by quantifying the extent that an episodic memory is distorted using a continuous measure of retrieval accuracy. In these experiments, participants encode stimuli that vary in similarity to a prototype, and memory for items that are least similar to the category prototype is biased toward it. For example, memory for color is biased toward canonical hues (Bae et al., 2015; Persaud & Hemmer, 2016), and size estimates for fruits and vegetables are biased by both their superordinate and subordinate mean sizes (Hemmer & Steyvers, 2009). The prototypes that influence memory retrieval in these experiments are either drawn from prior knowledge, (Hemmer & Steyvers, 2009; Persaud & Hemmer, 2016), or learned over the course of encoding (Allred et al., 2016; Duffy et al., 2010).

Notably, in all of these tasks, items whose features are most atypical, or whose features are farthest from their category prototype, are retrieved as closer to it, an effect that is interpreted as the consequence of integrating over item-specific memory and category knowledge. Yet, as described in the proceeding section, there is compelling evidence that memory for typical category members is more influenced by category information than atypical ones. In the current experiments, we assign typical and atypical category members to random spatial locations, which effectively uncouples stimulus typicality (i.e., typical or atypical bird) from the encoded and tested distance to the prototype (i.e., closer to or farther from to the location of most birds). This leads to a novel inquiry of how memory for the spatial location of a category member is biased by integrating information about its category typicality with information about the spatial location of its category.

#### **Typicality Modulates Encoding Strategies**

When might the use of prior category information to aid learning new knowledge change what is remembered about each imagelocation association? CAM posits that bias in memory arises at retrieval, as part of the integration of noisy event-specific estimates with category information. However, an alternative theory is that the presence of category information during learning changes how each event is encoded. For instance, categorization tasks are thought to orient attention to the stimulus dimensions that best differentiate the categories (Nosofsky, 1986). In the context of encoding new events, the awareness that category information is relevant for location may encourage relational encoding, which is thought to prioritize memory for distinctive or exception items by highlighting their dissimilarity with other encoded items (Hunt & McDaniel, 1993). This theory offers an interpretation of many findings of superior memory for items that are categorically isolated from surrounding items (von Restorff, 1933; Wallace, 1965) or that are inconsistent with or irrelevant to a salient task structure (Love et al., 2004; Sakamoto & Love, 2004).

A relational encoding strategy may benefit memory for inconsistent items by giving rise to their deeper, more elaborative processing (Craik & Tulving, 1975), as inconsistent items tend to receive more study time (G. R. Loftus & Mackworth, 1978; Stern et al., 1984), and changes to inconsistent stimuli are more likely to be noticed than changes to consistent ones (Friedman, 1979; Goodman, 1980; Heider, 1946; Schank & Abelson, 1977; Sentis & Burnstein, 1979). Taken together, this work suggests that an encoding task that encourages relational processing will shift the processing of atypical or distinctive items to be deeper, leading to more detailed memory, relative to typical items (Dunlosky et al., 2000; Hunt & Lamb, 2001).

In the context of the current experiments, the task of memorizing locations that largely map onto category membership may draw attention to the category membership of each item, encouraging relational encoding based on category. This in turn could promote deeper processing of atypical category members. On the other hand, as typical category members share more features with other members of their category, their processing may become more efficient, coming at a cost to memory for specific details. This reasoning gives rise to two predictions: First, we predict more bias in memory for typical over atypical category members. Second, in Experiment 3, we introduce a second test of memory for the exact images participants encoded, predicting better memory for this detail for atypical over typical category members.

#### **Overview of Experiments**

We aimed to quantify distortions in episodic memories due to the integration of prior category knowledge and new spatial information by examining how differences in category typicality bias new episodic memories. In a series of behavioral experiments, participants encoded and retrieved image-location associations on a two-dimensional (2D) grid (Figure 1A). The locations associated

with the images were determined by semantic relatedness ratings, such that most members of the same category (e.g., birds) were located near each other, but some typical and atypical category members were in random locations (Figure 1B-C). This configuration allows participants to learn that images from a certain category tend to be located in a certain area as they encode the locations of specific images. This general knowledge could only be learned and used if participants integrated learning of new imagespecific locations with their prior knowledge of animal and object categories. We used a continuous retrieval measure to disentangle biases driven by semantic knowledge from errors due to forgetting. Error and bias could vary independently, such that memory for an image could be biased toward or away from category neighbors at the same level of accuracy. We used these measures to quantify the influence of knowledge about a category's general location when retrieving the locations of each image.

We predicted that the use of this category knowledge would give rise to two effects: (a) retrieval of images located near members of the same category would be more accurate relative to those in random locations; and (b) for images located in random locations, retrieval of typical category members would be more biased toward their category's general location relative to atypical ones. Such a bias would reflect stronger associations between typical category members and their category neighbors. Finally, although we expected that category typicality would specifically influence bias in memory, we also explored potential differences in their accuracy.

We tested these predictions across four behavioral experiments (summarized in Table 1). In Experiment 1, we developed a novel, data-driven procedure to determine category membership and typicality for a set of animals and objects. In Experiment 2, we designed a conceptual replication using a different procedure for defining category typicality with validated and predetermined categories. As a preview, both experiments provided evidence in favor of the two predictions outlined above. In Experiments 3 and 4, we examined possible explanations for the observed differences in bias by category typicality. In Experiment 3, we asked whether the utility of category knowledge during encoding differentially influenced how typical and atypical category members processed, by assessing memory for their perceptual details. In Experiment 4, we developed a set of stimuli that could be organized by a nonsemantic property (i.e., arbitrary image color) to examine whether the observed biases in memory were driven by the visual similarity of the images rather than their category membership. Together these experiments show that, when learning new information that can be scaffolded by prior knowledge, integrating prior knowledge with new information can systematically enhance or distort memory for specific events, likely through shifts in encoding strategies.

### **Experiment 1**

In Experiment 1, participants encoded image-location associations, where the location of each image was either spatially consistent or spatially inconsistent with its category membership. This experiment was designed to test two predictions: (a) retrieval of an image's location would be more accurate if its encoded location was spatially consistent with its category membership, by being located near other category members; and (b) for images whose











*Note.* (A) Experiment 1 procedure. Participants encoded each image-location three times by dragging the image onto a red dot marking its location. After a 5-min break, participants retrieved the location of each image. (B) Image locations in the experimental group. In the control group, images were randomly assigned to the locations within each superordinate category (right and left side of the screen) such that they were no longer clustered by basic-level category. (C) Spatial consistency and category typicality for the "bird" category. Black indicates spatially "consistent" and green/bold font indicates spatially "inconsistent" images. All inconsistent images were either typical (italicized) or atypical (not italicized) category. (D) Example of retrieval measures for an image biased toward its category's cluster center. Solid red line indicates error. Solid blue line indicates unadjusted bias. Bias is quantified as an image's unadjusted bias as a proportion of error (solid blue line divided by red line). See the online article for the color version of this figure.

Figure 1

locations were not spatially consistent with their category membership, retrieval of a typical category member would be more biased toward its category's cluster center, relative to an atypical one.

#### Method

#### **Participants**

Participants completed all experiments on Amazon Mechanical Turk (AMT; see Supplemental Methods for eligibility requirements). Seventy participants were equally divided into the experimental group (N = 35; 15 females; mean age = 38.9; SD = 10.9; range 21–71 years) and the control group (N = 35; 16 females; mean age = 37.1; SD = 10.8; range 20–61 years). The sample size was chosen based on a power analysis estimating a medium effect size for a within-subjects comparison (d = .5,  $\alpha = .05$ , power = 0.8, two-sided, paired). This effect size is slightly more conservative than the median effect size across a variety of memory experiments (Morris & Fritz, 2013). Participants were paid \$5 upon completion of the experiment. The University of Pennsylvania IRB approved all consent procedures.

All participants who completed the experiment were included in the analyses reported here. We did not exclude participants based on performance, as performance-based exclusion criteria may have differed across the presented experiments due to differences in their difficulty or other task parameters, and we wished to keep recruiting procedures consistent across them.

#### Materials

Stimuli consisted of 70  $100 \times 100$  pixel color images of two superordinate semantic categories (animals and objects) on white backgrounds from the Bank of Standardized Stimuli (Brodeur et al., 2010). Piloting was conducted to choose the images in each condition to ensure that high levels of image recognition were matched across the two superordinate categories.

#### Image-Location Associations

Participants learned and retrieved associations between images and locations on a grid. The locations of images were determined by the semantic similarity of the concepts, as assessed by separate cohorts of participants who completed an odd-man out task (e.g., Connolly et al., 2007), a procedure where participants viewed three images and chose the image that was least similar to the other two. This procedure enabled us to produce a 2D array of image locations that closely resembled the relations among concepts in semantic space, clustering according to their category membership. This 2D mapping was the basis for the image locations in the memory experiment, as the exact coordinates derived from multidimensional scaling (MDS) of the semantic similarity data were used as image locations (with the exception of spatially inconsistent images; see next paragraph). The center of each category's cluster was derived by computing the mean x and y coordinates over all images belonging to that cluster. See the Supplemental Methods for more details about the odd-man-out procedure (Figure S6).

To disrupt the relationship between category membership and spatial location, we manipulated a subset of images such that their spatial locations no longer corresponded to their coordi-

nates from the MDS procedure (Figure 1C). To do this, the 20% most typical and 20% most atypical category members in a category cluster were assigned a new, random location, with the constraint that the location was closer to another cluster center than to its original cluster center. The remaining 60% of images were left in place, so that participants would be able to learn that different categories were generally associated with a particular spatial location. Thus, in total, 42 images were associated with locations that were consistent with their semantic similarity ("spatially consistent"), and 28 were displaced to a random location ("spatially inconsistent"). Of the 28 spatially inconsistent images, 14 were typical category members and 14 were atypical category members. Finally, the displays for animals and objects were arranged on the left and right sides of the screen, randomized for each participant. Thus, locations for most images were spatially consistent with semantic knowledge (e.g., several birds were located in a cluster), but some images were inconsistent (a few birds were located closer to other clusters of animals than to the other birds; see Figure 1B for an example display).

The above methods were applied to subjects in the experimental group. In a separate control group, the images paired with each location were shuffled separately for animals and objects such that both groups viewed the same locations, but the images did not cluster by category. In other words, the locations that had originally been associated with spatially (in)consistent or (a)typical category members were associated with random images, rendering these conditions meaningless. This control group is critical for two reasons. First, it provides a comparison group to identify potential effects of prior category knowledge on new episodic memories when the category knowledge is no longer useful for learning new image-location associations. Second, it ensures that any effects observed in the experimental group are driven by differences in the spatial consistency and typicality of images relative to their semantic category, rather than idiosyncrasies in the stimulus display (e.g., spatial clustering). For example, because images with inconsistent locations were located far from their category clusters, and because the clusters were dispersed across the display, more area of the grid is situated between an inconsistent location and its cluster center than in the opposite direction (away from its cluster center). This means that any response is more likely to be located somewhere in between an encoded location and its cluster center. It is for this reason that we included control groups in every experiment-their performance can be used as a baseline for the amount of bias that is induced by the constraints of the stimulus display.

#### **Procedure Overview**

The experiment comprised an encoding phase and a retrieval phase, separated by a 5-min break when participants were free to complete other tasks or Internet browsing (Figure 1A). Before beginning encoding, participants were presented with a white  $600 \times 1,200$  pixel rectangle with light gray gridlines spaced to form  $50 \times 50$  pixel grids. The size of the grid was derived from browser measurements of past AMT participants collected from pilot data to increase the likelihood that a participant could view the entire screen without zooming out. Participants were instructed

to adjust the size of their browser, without zooming, so they could see the entire grid. The experiment would not advance until the entire grid was visible.

#### Encoding

On each trial, participants viewed an image beneath the grid and a red dot on the grid corresponding to that image's location. They were instructed to drag each image onto the dot, to click the mouse button or press the "enter" key once the image was positioned over the dot, and to try to remember each image's location for a later memory test. After this movement, the participant immediately advanced to the next trial. Trials were self-paced, and participants were unable to proceed to the next trial if they had not moved the image directly over the red dot. To ensure sufficient memory for the locations, all trials were presented three times, in separate blocks. The order of trials was randomized for each participant, separately for each block. The encoding instructions were followed by two practice trials to familiarize participants with the task before beginning encoding.

#### Retrieval

The timing and task were identical to those from the encoding phase, but with no red dot marking the location of the image. Participants were instructed to drag the image to its associated location and to make their best guess if they had forgotten it. They did not receive feedback during this task. The trial order was randomized for each participant.

#### Statistical Analyses

We developed two measures to quantify memory accuracy (error) and the influence of category knowledge (bias) for each image (Figure 1D). Error was defined as the Euclidian distance between an image's encoded location and its retrieved location, where greater values indicate worse memory, and perfect memory would correspond to an error of 0. Bias was defined as the proportion of total error in the direction of an image's category cluster. To compute this, we first found each image's unadjusted bias by subtracting the Euclidean difference between its encoded location and its cluster center from the Euclidean difference between its retrieved location and cluster center. Then, we divided this unadjusted bias by the amount of error for the image. Thus, a bias score between 0 and 1 indicates that retrieval is biased toward the image's cluster center, and a score between 0 and -1 indicates that retrieval was biased away from the cluster center. In other words, error refers to the *magnitude* of error for each image, and bias represents the *direction* of error, or the proportion of error that is closer to or farther from its category center. These two measures are statistically independent, such that two memories with the same amount of error could differ in whether their retrieval was biased toward or away from their respective cluster centers.

Both measures were averaged across trials, by spatial consistency (consistent vs. inconsistent), and then by category typicality (atypical vs. typical). Effects on average error and bias were assessed with two-tailed paired t tests and repeated measures ANOVAs. We report effect sizes for all significant effects, including partial  $\eta^2$  for main effects or interactions of ANOVAs and Cohen's d for effect sizes of within-subject comparisons. The equation used to calculate Cohen's d accounted for the high correlations across these within-subject conditions (Goulet-Pelletier & Cousineau, 2018; Lakens, 2013):

$$d = \frac{M_2 - M_1}{\sqrt{S_1^2 + S_2^2 - 2rS_1S_2}}$$

 $M_1$  and  $S_1$  are the mean and standard deviation of one condition,  $M_2$  and  $S_2$  are the mean and standard deviation of the other condition, and *r* is the correlation between the two. Bootstrapped 95% confidence intervals for these effect sizes were generated as well. Considering the correlation between conditions results in effect sizes that are more representative of the reported withinsubject statistical tests and are comparable across the four experiments reported here, but it is important to note that such effect sizes are inflated compared with effect sizes from betweensubjects designs and thus may be less generalizable to other experiments in the literature (Lakens, 2013).

#### Software

Stimuli were presented using customized scripts written in HTML and JavaScript. MturkR was used to interface with AMT to post experiments, retrieve data, and pay participants (Leeper, 2017). All statistical analyses were conducted in R (R Core Team, 2018). *ImerTest* was used to provide statistics of main effects and interactions for mixed-effects models (Kuznetsova et al., 2017), *emmeans* was used to compute estimated marginal means to provide contrasts of their simple effects (Lenth, 2018), and *bootES* was used to generate bootstrapped confidence intervals around their estimated effect sizes (Kirby & Gerlanc, 2013). All data figures were generated with *ggplot2* (Wickham, 2016).

# Results

# Spatial Consistency

We first tested the prediction that location memory would be more accurate for images that were near their category center relative to images in random locations (Figure 2A). As a reminder, spatially inconsistent images comprise both typical and atypical category members, while spatially consistent images comprise category members of intermediate typicality. We computed a 2 (group: experimental, control)  $\times$  2 (spatial consistency: consistent, inconsistent) ANOVA which revealed a main effect of spatial consistency, F(1, 68) = 26.96, p < .001,  $\eta_p^2 = .28$ , no main effect of group, F(1, 68) = 0.75, p = .39,  $\eta_p^2$  = .28, and a significant interaction between group and spatial consistency, F(1, 68) = 18.10, p < .001,  $\eta_p^2 = .21$ . This interaction reflected more error for spatially inconsistent images relative to consistent images in the experimental group, t(34) = -7.31, p < .001, d = -1.24, but not in the control group, t(34) = -0.61, p = .54, d = -0.10, where images did not cluster by category (Figure 2B). Location memory was more accurate for images that were located near their category centers relative to those that were in random locations, suggesting that over the course of the experiment, participants learned that different categories were likely to be located in certain

# Figure 2 Experiment 1 Predictions and Results



Note. (A) We predicted that images whose locations were spatially consistent with their category (e.g., located near other category members) would be remembered more accurately than images in random locations. Gray circle indicates the center of a category cluster. Arrows indicate the encoded and retrieved locations of images from that category. Shorter arrows signify more accurate memory. Note that "spatially inconsistent" locations comprise the 20% most typical and 20% most atypical category members, while "spatially consistent" locations comprise images of intermediate typicality. (B) Average error as a function of spatial consistency, where greater values indicate larger errors in location memory. Scrambled "consistent" and "inconsistent" conditions in the control group indicate error for images that were randomly assigned to the spatially consistent and inconsistent locations in the experimental group. (C) Of all spatially inconsistent images, we predicted that typical category members would be retrieved closer to their category center relative to atypical members. (D) Average bias by category typicality and group, where >0 indicates retrieval closer to the category center than originally encoded and <0 indicates retrieval away (B, D). Scrambled "typical" and "atypical" conditions in the control group indicate images randomly assigned to the locations of typical and atypical category members in the experimental group. Gray lines indicate participants. Error bars indicate standard error of the mean (SEM) across participants. \* p < .01. \*\*\*\* p < .001. See the online article for the color version of this figure.

parts of the grid, and the use of this knowledge aided memory for images whose locations matched this spatial organization.

# Category Typicality

As a reminder, in cases where the location of an image was not aligned with its category membership, we predicted that retrieval of typical category members would be biased toward their category center (Figure 2C). To test this prediction, we conducted a 2 (group: experimental, control) × 2 (category typicality: typical, atypical) ANOVA with bias of spatially inconsistent images as the dependent variable. We found no reliable main effect of typicality, F(1, 68) = 3.57, p = .06,  $\eta_p^2 =$ .05, or group, F(1, 68) = 2.06, p = .16,  $\eta_p^2 = .13$ , but a reliable interaction, F(1, 68) = 11.69, p = .001,  $\eta_p^2 = .15$  (Figure 2D). This interaction reflected greater bias for typical category members relative to atypical category members in the experimental group, t(34) = 3.52, p = .001, d = 0.60, but not the control group, t(34) = -1.16, p = .25, d = -0.20. As predicted, relative to atypical category members, typical category members were more biased, with a larger proportion of their error in the direction of their category center.

We were concerned that these effects might have arisen simply because typical images were placed more centrally in the grid, which happens to be close to most category clusters, while atypical images were placed at edges of the grid far from the clusters. To test this possibility, we computed each image's average bias toward the two other clusters in the category. A 2 (group: experimental, control)  $\times$  2 (category typicality: typical, atypical) ANOVA found no main effects or interaction (all F < 0.07, all p > .80). This suggests that participants' retrieval of typical category members was biased specifically toward the cluster center corresponding to their category. Furthermore, as landmarks often influence spatial memory (e.g., Huttenlocher et al., 1991), we also aimed to determine whether proximity to the edge of the grid or to the boundary between superordinate categories influenced bias or error. To do this, we divided the grid into locations that were near to or far from landmarks, where locations near landmarks were within 100 pixels of the edge of the grid or the vertical line dividing the grid into right and left sides. We did find that error and bias differed by whether images were located near landmarks, but critically, these effects did not interact with the observed effects of error by spatial consistency and bias by category typicality. See the Supplemental Results for Experiment 1 for more details.

After confirming that bias in memory differed according to an image's category typicality, we next explored whether category typicality influenced memory accuracy (Figure S1). To do this, we entered the average error of spatially inconsistent images into a 2 (group: experimental, control)  $\times$  2 (category typicality: typical, atypical) ANOVA. This revealed no main effect of group, F(1,68) = 0.02, p = .88,  $\eta_p^2 = 0.004$ , or of typicality, F(1, 68) = 0.79,  $p = .38, \eta_p^2 = .01$ , but a reliable interaction, F(1, 68) = 4.44, p =.04,  $\eta_p^2 = .06$ . This interaction reflected greater error for typical images relative to atypical images in the experimental group, t(34) = 2.34, p = .03, d = .40, but not in the control group, t(34) = -0.79, p = .43, d = -.13, suggesting that location memory for atypical category members was more accurate relative to location memory for typical category members. Taken together, this suggests that category typicality influenced both the magnitude and direction of error in location memory for images whose spatial locations did not correspond to their category membership.

#### Discussion

In Experiment 1, we found that participants' retrieval was more accurate for images that were spatially consistent with their category membership, in line with prior observations that new memories are enhanced if they are consistent with prior knowledge. Our findings are different from this past work, however, because new memories were enhanced if they were consistent with newly learned information about category locations. This demonstrated that participants were able to learn that the locations of many images aligned with their category membership and use this new knowledge to aid retrieval of images whose locations were consistent with this organization.

Furthermore, we observed biases in location memory that reflected the use of prior category knowledge when learning about the locations of each category. First, regardless of category membership, images that were located randomly were retrieved closer to their category's cluster center relative to their encoded location. Already, this suggests that participants' reliance on their category knowledge distorted memory for images whose locations did not map on the spatial organization that mirrored their category membership. Critically, we also found variation in the extent to which these memories were distorted. When sorting the spatially inconsistent images by category membership, typical category members were more biased toward their category neighbors relative to atypical ones. This finding provides novel evidence that variations in the organization of semantic knowledge, like category membership and typicality, can govern the extent of distortion in new episodic memories.

### **Experiment 2**

Category membership and typicality in Experiment 1 were derived using a data-driven procedure, based on semantic judgments provided by a separate cohort of participants. We developed Experiment 2 as a conceptual replication with three goals. First, we aimed to investigate whether we would observe the same effects using conditions that were derived from a more validated procedure for identifying typical and atypical category members. Second, we aimed to rule out the possibility that memory was more accurate for spatially consistent images because their spatial locations were more densely clustered, increasing the likelihood of guessing the correct location. Replicating the effects observed in Experiment 1 with these changes would confirm that these effects generalize across different task parameters and an expanded set of images.

Third, we added a confidence measure to understand how the strength of memory for each image location may modulate its influence by category knowledge. Past work considering memory as a Bayesian reconstruction process has shown that less confident memories are prone to more error and are more biased by category information, suggesting that weaker memories for specific events are more likely to draw on other sources of memory (Brady et al., 2018; Persaud & Hemmer, 2016). We predicted that, if participants drew on a combination of memory of specific locations and knowledge of category locations, they would rely more on category knowledge when memories for specific locations were weak. This would result in (a) a larger difference in accuracy for spatially inconsistent versus consistent images for low-confident memories; and (b) more biased retrieval, where low-confident images would be retrieved closer to their category center (Brady et al., 2018). As we did not have strong predictions about whether confidence would differentially relate to the observed typicality effects, we conducted an exploratory analysis of their interaction.

# Method

# **Participants**

Seventy participants took part in Experiment 2, half in the experimental group (N = 35; 14 females; mean age = 38.8; SD = 10.4; range = 22–70 years) and half in the control group (N = 35; 16 females; mean age = 38.3; SD = 11.8; range = 24–72 years). Participants were recruited via AMT with the same eligibility criteria as in Experiment 1. Participants were paid \$10 upon completion of the experiment. As the smallest effect size in Experiment 1 (greater bias for typical over atypical category members in the experimental group, d = 0.60) was larger than the estimated effect size used to plan our sample size for Experiment 1, we chose to recruit the same number of participants for Experiment 2 for consistency across the experiments.

### **Materials**

Stimuli comprised 160  $100 \times 100$  pixel color images on white backgrounds (80 animals, 80 objects). These two superordinate categories were each divided into four categories with 20 images each: birds, insects, sea creatures, mammals, clothes, furniture, kitchen, and office. The categories were selected from prior studies of categorization norms (Deyne et al., 2008; Uyeda & Mandler, 1980). Category typicality was determined with a list ranking procedure completed by a separate group of participants. See Supplemental Methods and Figure S7 for details of this procedure and the resulting rankings.

To generate images' locations, the memory grid was divided into halves with animals on one side, and objects on the other. On each side, all images were spaced uniformly apart, resulting in an even distribution of images across the grid. Each side's locations were divided into four quadrants, and the four categories were randomly assigned to different quadrants.

Like in Experiment 1, the spatial locations of a subset of images were manipulated such that their locations no longer were consistent with semantic knowledge (Figure 3A). To do this, the 15% most typical and 15% most atypical category members were swapped with the typical and atypical members of other categories such that each quadrant had an equal number of typical and atypical images from the other three categories. The remaining 70% of images were randomly assigned to locations within their category's quadrant. In total, 112 images were in locations that were spatially consistent with their category membership ("consistent"), and 48 were associated with a random location ("inconsistent"). Of the 48 inconsistent images, 24 were typical and 24 were atypical category members. The sides (animals on right or left) and quadrants were randomized for each participant.

#### Procedure

The timing, task, instructions, and statistical analyses were identical to those of Experiment 1, with the exception of an added confidence measure. After each retrieval trial, participants chose from four options to report the confidence in their location memory: very confident, somewhat confident, guessed, and forgot item. Clicking on an option automatically advanced the participant to the next trial.

# Statistical Analyses

Analyses of error by spatial consistency and bias by category typicality were computed identically as in Experiment 1. In addition, error and bias were analyzed as a function of confidence. The options were not consistently used; as an example, some participants never chose "forgot item" (see Supplementary Tables 1 and 2 for the percentage of reported confidence responses divided by spatial consistency and category typicality). Because there were unequal amounts of data in each confidence bin, linear mixedeffects models were used to investigate how confidence related to error and bias. Participant intercepts and slope terms for each included predictor variable were modeled as random effects. The significance of a given contrast was obtained using Satterthwaite approximate degrees of freedom, resulting in F or t statistics and corresponding p values. Estimated marginal means (EMMs) were computed to test simple effects. These were corrected for multiple comparisons with a Bonferroni correction, depending on the number of tests conducted. Where this correction is applied, the raw pvalues are reported alongside the corrected  $\alpha$  and number of tests conducted. Figures 3C and 3E, which visualize the results of mixed-effects models, display averages across trials within participants for each condition for easier interpretation. To visualize effect sizes, we computed Cohen's d for all pairwise contrasts of participants' average error and bias separately at each level of confidence (Figure S4).

# Results

# Spatial Consistency

Like in Experiment 1, we first computed a 2 (group: experimental, control) × 2 (spatial consistency: consistent, inconsistent) ANOVA to test whether spatial consistency with category knowledge influenced the accuracy of location memory. This revealed a main effect of consistency, F(1, 68) = 38.63, p < .001,  $\eta_p^2 = .36$ , and no main effect of group, F(1, 68) = 0.01, p = .91,  $\eta_p^2 = .008$ . There was a significant interaction between group and consistency, F(1, 68) = 35.33, p < .001,  $\eta_p^2 = .34$ , explained by greater error for spatially inconsistent images relative to spatially consistent images in the experimental group, t(34) = -7.35, p < .001, d = -1.24, but not in the control group, t(34) = -0.24, p = .81, d = -0.04 (Figure 3B). Replicating the observations in Experiment 1, retrieval was more accurate for images located close to their category center.

We next sought to investigate how error was modulated by participants' reported confidence. Focusing on the experimental group, we entered consistency and confidence of each trial into a mixed-effects model with error as the dependent measure. We found a strong interaction between spatial consistency and confidence, F(3, 1522.36) = 3.14, p = .02, in addition to significant main effects of consistency, F(1, 49.27) = 38.77, p < .001, and confidence, F(3, 46.90) = 83.18, p < .001 (Figure 3C). The main effect of confidence reflected less error for more confident responses relative to less confident responses for all pairwise comparisons of confidence responses (all t < -3.5, all p < .002,  $\alpha =$ .008 with Bonferroni correction for six tests). Pairwise tests of error by consistency for each confidence response revealed that the difference in error by consistency was greater for less confident memories (very confident: t(398.18) = -0.52, p = .60; somewhat confident: t(100.17) = -6.13, p < .001; guessed: t(119.95) =-5.58, p < .001; forgot item: t(663.30) = -3.70, p < .001;  $\alpha =$ .0125 with Bonferroni correction for four tests), confirming our prediction of a greater use of category knowledge when retrieving weaker memories relative to strong ones.

## Category Typicality

As in Experiment 1, we examined how category typicality modulated bias in memory, again predicting that retrieval of typical category members would be more biased toward their category center than retrieval of atypical category members. We conducted the same 2 (group: experimental, control) × 2 (category typicality: typical, atypical) ANOVA of bias for spatially inconsistent images. This analysis revealed a significant interaction between group and typicality, F(1, 68) = 4.20, p = .04,  $\eta_p^2 = .06$ , in addition to a main effect of group, F(1, 68) = 6.61, p = .01,  $\eta_p^2 =$ .28, and no main effect of typicality, F(1, 68) = 0.78, p = .38,  $\eta_p^2 = .01$  (Figure 3D). As in Experiment 1, the interaction reflected more bias for typical category members relative to atypical category members in the experimental group, t(34) = 2.14, p = .04, d = .36, but not the control group, t(34) = -0.80, p = .43, d = -0.14.

We also investigated how confidence modulated bias with a mixed-effects model of confidence and typicality as predictors, focusing only on the experimental group (Figure 3E). This analysis revealed a main effect of confidence, F(3, 48.69) = 10.04, p < 1000



Α

Corkscrew Manatee Flamingo Swordfish Great White Mussel Frying Pan Couch Cup litten Strainer Pot Rolling Pi Overalls MagnifyingMeasuring Spoor Int Rolling Pin Swallow Caterpillar Killer Whal Eagle Duck Chicken Seal Starfish Rooster Cockroach Clownfish Rhinoceros Lobster leanGlaseasuring CupBed Pen Fondue Pot Ladle Pelican Cravon Plate Pencil Swan Sea Turtle Cardinal Dresser bula Bib Sea Lion Coat Rack Peeler Crab HippopotamusHeron Shrimp Fork es Spatula Seagull Cow Penguin Sauid Potato MasherBowl Owl Stingray nmerhe Parrot Girafie Bison Kev Beetle BathTub Desk Envelope Ladvbug Butterfly Drag Spoon Thumbtack Ostrich Sink Lion Fea Kettle Salt ShakeGarbage CarClipboard Keyboard Wasp Fox Toucan Bluejay Lamp Si Night Stand T-Shirt Kangaro Stool Таре Paintbrush Skunk Daddy Long Legamel Bug Racoon Ruler Zebra Elephant Shoe Paper Clip Binder Stapler Lock Elee Dolphin Tiger " Table Curtain Chalk Mouse Praying Ma Spider Moose Pig Chest Lightbulb Knife Sheep Narwha Bookshelf Termite Wolt Oyster HighlighterBike H Imet Pillow Moth Pigeon . .lellvfish Chandelier Chair Cheetah



*Note.* (A) Image locations. (B) Average error by spatial consistency, where greater values indicate less accurate location memory. (C) Average error by spatial consistency and reported confidence in the experimental group. Statistics reflect the simple effects of a mixed-effects model. (D) Average bias of spatially inconsistent images by typicality and group, where >0 indicates that trials were retrieved closer to their category cluster than originally encoded and <0 indicates retrieval away. (E) Average bias by typicality and reported confidence in the experimental group. Statistics reflect mixed-effects models (B–E). Gray lines indicate participants. Gray dots signify participants with no responses in the other bin (e.g., a dot in Forgot Item for consistent images indicates that the participant did not use "forgot item" for any inconsistent images). Error bars indicate standard error of the mean (*SEM*) across participants. Scrambled conditions in the control group indicate retrieval of the images at locations to which spatially (in)consistent and (a)typical images had been assigned in the experimental group; these locations were randomly assigned images in the control group.

\* p < .05. \*\*\* p < .001. See the online article for the color version of this figure.

.001. There was no reliable effect of typicality, F(1, 75.45) = 0.22, p = .64, and no significant interaction between typicality and confidence, F(3, 1033.28) = 2.09, p = .10. To understand the main effect of confidence, we computed comparisons across confidence responses, collapsed over typicality ( $\alpha = .008$ , Bonferroni

correction for six tests). We found that "very confident" responses were less biased toward their cluster center relative to "somewhat confident", t(23.29) = -3.87, p < .001, and relative to "guessed", t(28.46) = -3.0, p = .006. Interestingly, "forgot item" responses were less biased relative to "guessed" responses, t(18.66) = 2.76,

Image Size p = .01, and relative to "somewhat confident" responses, t(22.61) = 2.89, p = .009, but these comparisons do not survive correction for multiple comparisons.

We next aimed to replicate the influence of category knowledge on the accuracy of spatially inconsistent images observed in Experiment 1. We conducted the same 2 (group: experimental, control) × 2 (category typicality: typical, atypical) ANOVA of average error for spatially inconsistent images (Figure S1). This revealed no effect of group, F(1, 68) = 1.24, p = .27,  $\eta_p^2 = .19$ , and a reliable main effect of typicality, F(1, 68) = 9.65, p = .003,  $\eta_p^2 = 0.12$ . This was qualified by an interaction between typicality and group, F(1, 68) = 5.77, p = .02,  $\eta_p^2 = .08$ , such that there was greater error for typical category members relative to atypical ones in the experimental group, t(34) = 3.44, p = .002, d = .58, but not in the control group, t(34) = 0.59, p = .56, d = .10. As in Experiment 1, memory for the location of typical category members.

# Discussion

In Experiment 2, we replicated and extended the findings of Experiment 1. Across both experiments, participants were able to learn that images of a certain category were likely to appear in the same area of the grid, and this new learning affected their retrieval of specific image locations. First, memory was more accurate if the images spatially clustered near others from the same category. For images that were randomly located far from their category cluster, participants made systematic errors: memory for typical category members was less accurate than for atypical ones, and a larger proportion of their error was in the direction of their category's cluster center. These results were observed in both experiments despite differences in the numbers and types of categories, method of determining typicality, and mapping between category membership and spatial location. These observations provide strong evidence for the theory that episodic retrieval is supported by the integration of multiple sources of information: memory for specific image-location associations, prior knowledge of categories, and newly learned mapping of categories to spatial locations. This integration can either benefit or harm memory for individual images, depending on whether these sources are in conflict.

The introduction of a confidence measure provided further support for this theory. First, we found that for low-confident responses, error was more likely to be modulated by spatial consistency such that images located far from their cluster center were remembered less accurately than those located near their cluster center. However, as confidence in location memory increased, there was a less of a difference in error by consistency. This suggests that to retrieve stronger episodic memories, there is less of a need to rely on other information, like knowledge of the clustering of the images by category. However, it is also possible that the reverse is true-participants may have reported lower confidence because they knowingly based their retrieval of a specific location based on its category's general location. This may be particularly likely because confidence reports were collected directly after retrieval. The current experiment cannot adjudicate between these two interpretations, but both are consistent with the notion that retrieval of weaker memories is more affected by knowledge of category locations.

When focusing on spatially inconsistent images, we found that low-confident memories were more biased toward category neighbors relative to high-confident memories, again consistent with the notion that less strong episodic memories will be more distorted by other sources of knowledge (Brady et al., 2018). The finding that memory was less biased for images which were forgotten entirely is puzzling; this effect should be interpreted with caution as this option was used the least frequently (see Table S2) and consequently is the most underpowered and most variable condition in this analysis.

#### **Experiment 3**

Experiments 1 and 2 suggest that the way that prior knowledge is organized can shape how new episodic memories are formed and retrieved in the context of learning new information. In particular, we find that prior knowledge biases new memories differently based on their category typicality such that typical category members are more prone to bias by the location of other category members. Here, we probe whether this difference in bias by typicality is a consequence of participants' orientation toward category membership during encoding. As highlighted in the Introduction, it is well known that different encoding strategies give rise to different patterns of memory, where tasks that encourage relational encoding result in prioritized memory for distinctive or exception items by highlighting their dissimilarity with other encoded items (Bejjani & Egner, 2019; Goodman, 1980; Hunt & McDaniel, 1993; Love et al., 2004; Sakamoto & Love, 2004). Furthermore, a salient organizational structure can change which features of an item are more deeply encoded. Evidence for this idea can be found in experiments where participants learned that certain facial attributes dictated a face's spatial location. Location memory was better for faces with these attributes, but memory for other facial attributes was impaired, suggesting that the presence of a guiding organization during encoding aided memory for features that adhered to the organization at the cost of memory for idiosyncratic features (Sweegers et al., 2015; Sweegers & Talamini, 2014). If typical category members in the present experiments are thought of as more adherent to the organizational properties that determine category membership, and thus are less deeply processed, we predicted that memory for their idiosyncratic features may be impaired relative to that of atypical category members, thus providing an explanation for the increased bias in location memory for typical category members observed in Experiments 1 and 2.

We tested this prediction by assessing whether memory for a different feature of the stimuli was modulated by category typicality. To do this, we adapted the procedure in Experiment 1 by adding an exemplar memory test. During retrieval, participants were tested for their exemplar memory by choosing between the encoded image and a highly similar perceptual lure. They then dragged the chosen image to its encoded location. We predicted that exemplar memory for typical category members would be worse than for atypical members in the experimental group, where category membership was related to most images' locations and thus was a salient feature for participants to learn during encoding. We expected this influence of category typicality on exemplar memory to be stronger in the experimental group than in the control group, where participants encoded the same images but in random locations.

Whether typicality would influence exemplar memory in the control group was less clear. Relative to typical items, atypical or distinctive items are more often recalled and more likely to be successfully recognized across many stimulus classes and protocols (e.g., von Restorff, 1933; Wallace, 1965). But such effects are moderated by the typicality of surrounding list items, because when embedded in word lists of moderately typical animals and objects, typical category members are more likely to be recalled over atypical ones (Greenberg & Bjorklund, 1981; Schmidt, 1996). Moreover, these paradigms do not test recognition of the specific exemplars encoded, and it is less clear how memory for such perceptual details may be affected by category membership. Nevertheless, regardless of what we observed in the control group, we predicted a stronger modulation of exemplar memory by typicality in the experimental group that would reflect a greater difference in processing of the images due to their spatial organization.

We also conducted exploratory analyses to understand how exemplar memory interacted with location memory. Like exemplar memory, the location associated with each image can be thought of as an idiosyncratic feature of memory for the animals and objects, because the locations are newly learned and specific to the context of the experiment. Importantly, for spatially inconsistent images, this feature is in direct conflict with the organizational scheme guiding most image locations. Already, our findings in Experiments 1 and 2 are consistent with this notion when considering the location of the spatially inconsistent images as an idiosyncratic detail. Memory for the locations of typical category members was less accurate and more biased than memory for locations of atypical category members, perhaps because their locations were less deeply encoded. Thus, we predicted that poor exemplar memory would be related to even stronger bias for typical category members, as memory for both features may be correlated within each image-location association.

# Method

## **Participants**

In Experiment 3, 230 participants were divided into the experimental group (N = 115; 44 females; mean age = 35.3; SD = 8.4; range = 22–63 years) and the control group (N = 115; 55 females; mean age = 36.7; SD = 10.0; range = 19–67 years). Participants were recruited via AMT with the same procedures as in Experiments 1 and 2. To determine the sample size, we recruited 35 participants in the control group and used their performance on the exemplar memory test to estimate the number of participants needed to sufficiently power analyses of error and bias as a function of exemplar memory accuracy (see Supplemental Methods for more details of this procedure). Participants were paid \$5 upon completion of the experiment.

#### Materials

Experiment 3 was designed to be as similar as possible to Experiment 1, while adding exemplar memory and confidence measures during retrieval. Thus, encoded stimuli consisted of 70 color images on white backgrounds (35 animals, 35 common objects) and these stimuli were shown during retrieval along with

35 animal lures and 35 object lures. The images were taken from a variety of public data sets (Brodeur et al., 2010; Konkle & Caramazza, 2013; Russo et al., 2018; Stark et al., 2013). The 70 animals and objects presented during encoding corresponded to those in Experiment 1, but the images representing each animal and object were replaced by images drawn from the data sets published by Stark et al. (2013), which feature perceptually matched pairs of images commonly used in behavioral studies examining pattern separation (Kirwan et al., 2007). If an image from Experiment 1 did not have a corresponding pair of images in this database, a second image was found from a different database or from the Internet and matched with the one used in Experiment 1. Analyses of same 35 control participants used to determine the sample size demonstrated that exemplar memory for a given image was not at ceiling or at floor.

#### Procedure

The timing, task, and instructions were identical to that of Experiment 1, except for the retrieval task. On each retrieval trial, a similar lure appeared next to the encoded image (right/left presentation randomized). Participants were instructed to choose the image they had encoded, drag it to its location, and confirm their answer by clicking continue or pressing the space bar (Figure 4A). Then, participants completed two confidence measures. First, they assessed confidence for the image they chose (very confident, somewhat confident, completely guessed). Next, they assessed confidence of their retrieved location (very confident, somewhat confident, completely guessed). See the Supplemental Results for analyses of exemplar memory, error (Figure S2B), and bias (Figure S3B) by confidence.

#### Statistical Analyses

Average exemplar memory accuracy by category typicality was analyzed with planned two-tailed *t* tests as well as with a 2 (group: experimental, control)  $\times$  2 (typicality: typical, atypical) ANOVA. Analyses of error and bias as a function of exemplar memory were performed using mixed-effects models, identical to how effects of confidence were analyzed in Experiment 2. Figures 4C and 4D, which visualize the results of mixed-effects models, display averages across trials within participants for each condition for easier interpretation. See Supplementary Figure S4 for effect sizes of average error and bias by exemplar memory.

#### Results

#### **Exemplar Memory**

We first tested the prediction that the image details of atypical category members would be more deeply encoded than those of typical category members. To do this, we compared exemplar memory as a function of category typicality. As a reminder, we predicted that a difference in exemplar memory by typicality would be present in the experimental group, where category membership was related to most images' locations and thus was a salient feature for participants to learn and use. We quantified exemplar memory as the proportion of trials in which participants chose the correct image. Note that in this analysis, in contrast to all other analyses conducted up to this point, typicality in the control group refers to the true typicality of the category members, not of



*Note.* (A) Retrieval instructions and display. (B) Performance on the exemplar memory test by group and category typicality. (C) Average error in the experimental group, by spatial consistency and by performance on the exemplar memory test (correct vs. incorrect). Statistics reflect results of trial-level mixed-effects models. (D) Average bias by typicality and performance on the exemplar memory test in the experimental group, where values >0 indicate that trials were retrieved closer to neighbors than originally encoded and values <0 indicate retrieval away from neighbors (B–D). Statistics reflect simple effects that survive a Bonferroni correction for four comparisons ( $\alpha = .0125$ ). Gray lines indicate participants. Gray dots signify participants with no responses in the other bin (e.g., a dot in Exemplar Incorrect for typical category members). Error bars indicate *SEM*. \*\* p < .01. \*\*\* p < .001. See the online article for the color version of this figure.

the images that were located in the positions assigned to typical category members in the experimental group.

As we predicted, we found that atypical category members were more accurately recognized than typical ones in the experimental group, t(114) = 2.96,  $p = .004 \ d = .28$ , but not in the control group, t(114) = 0.94, p = .35, d = 0.09 (Figure 4B). When comparing across groups, the experimental group exhibited better exemplar memory for atypical category members, t(228) = 3.10, p = .002, d = 0.41, and to a lesser extent for typical ones, t(228) =1.95, p = .05, d = 0.26. However, a 2 (group: experimental, control)  $\times$  2 (typicality: typical, atypical) ANOVA did not reveal a significant interaction between group and typicality, F(1, 228) =1.95, p = .16,  $\eta_p^2 = .008$ , so differences in exemplar memory by typicality across the groups should be interpreted with caution. This ANOVA also revealed a significant main effect of group, F(1, 228) =8.07, p = .005,  $\eta_p^2 = .12$ , and of typicality, F(1, 228) = 7.48, p = .007,  $\eta_p^2 = .03$ . In summary, when images were presented in a context where category membership was relevant to the location memory task, participants exhibited better memory for the image details of atypical category members relative to typical ones.

# Spatial Consistency

Before considering how exemplar memory influenced location memory accuracy, we first confirmed that retrieval of spatially consistent images was more accurate than retrieval of images in random locations, replicating our observations in Experiments 1 and 2 (see Supplemental Results). We next explored whether exemplar memory and consistency interacted to influence accuracy, focusing on the experimental group. To do this, we computed a mixed-effects model with error as the dependent variable and spatial consistency, exemplar memory accuracy, and their interaction as independent variables (Figure 4C). Exemplar memory reliably related to error, F(1, 123.6) = 20.57, p < .001. There was no reliable interaction between exemplar memory and consistency, F(1, 6178.8) = 3.55, p = .06. In post hoc comparisons ( $\alpha = .0125$ when Bonferroni-corrected for four tests), we found that for spatially inconsistent images, error was reliably greater when exemplar memory was incorrect relative to when exemplar memory was intact, t(430.92) = -4.32, p < .001, while for spatially consistent images, this difference did not survive correction for multiple comparisons t(234.77) = -2.35, p = .02. At the same time, there was greater error for spatially inconsistent relative to consistent images regardless of exemplar memory accuracy (exemplar memory correct: t(187.92) = -8.14, p < .001, exemplar memory incorrect: t(955.07) = -7.10, p < .001). In sum, exemplar memory accuracy was related to location memory only for spatially inconsistent images, with worse performance on the exemplar test relating to less accuracy in location memory.

# Category Typicality

After replicating the observation from Experiments 1 and 2 that typical category members were retrieved closer to their category center than atypical category members (see Supplemental Results), we next tested whether exemplar memory and typicality interacted to influence bias in location memory (Figure 4D). We predicted that typical category members with poor exemplar memory would be more biased than ones with intact exemplar memory. We also conducted exploratory comparisons of bias by typicality separately for correct and incorrect exemplar memory ( $\alpha = .0125$ , four tests). Contrary to our prediction, we found no reliable difference in bias by exemplar memory accuracy, neither for typical category members, t(306.43) = 0.38, p = .70, nor atypical category members t(316.27) = -0.86, p = .39. When considering images that were correctly remembered, retrieval was more biased for typical relative to atypical category members, t(192.84) = 4.52, p < .001. In contrast, there was no reliable difference of typicality for images with incorrect exemplar memory, t(830.02) = 1.64, p = .10. However, we found no interaction between exemplar memory and typicality, F(1, 3069.36) = 0.87, p = .35, so differences in bias by category typicality should be interpreted with caution. Finally, there was no reliable main effect of exemplar memory, F(1,101.27) = 0.12, p = .73. In sum, when participants remembered the details of the images, typical images were retrieved closer to their category center relative to atypical images, but when participants forgot the exemplar, this difference in bias by category typicality was attenuated.

#### Discussion

In Experiment 3, we sought to understand whether differences in depth of processing led to differences in how participants remembered typical and atypical category members. We found that, when category information was salient and useful during encoding, there was better memory for the image details of atypical category members over typical ones. This suggests that the idiosyncrasies of typical category members are more poorly remembered than those of atypical category members, consistent with the notion that typical category members may be processed less deeply and more efficiently than atypical ones. Critically, this difference in processing occurred only in an encoding context where category information was made more salient by the associated spatial configuration. In the control group, where category information did not dictate the locations of images, there was no difference in exemplar memory for typical and atypical category members. While this finding departs from prior work showing that prior category knowledge gives rise to differences in memory for typical and atypical items (e.g., Greenberg & Bjorklund, 1981; Hunt & Lamb, 2001; Whitney et al., 1983; Schmidt, 1996), it may be that memory for perceptual details is less influenced by category typicality without a task that draws attention to category membership. Note that the lack of a group by exemplar memory interaction prevents us from strongly interpreting these differences in exemplar memory across the experimental and control groups.

We also replicated the observation in Experiments 1 and 2 that images located in random locations were less accurately remembered relative to images located close to their category center. Furthermore, this difference in memory accuracy was magnified for images with incorrect exemplar memory, although the interaction between spatial consistency and exemplar memory was not statistically reliable. Nevertheless, this gives rise to an intriguing possibility that for spatially inconsistent images, exemplar accuracy is a marker for the fidelity of each image-location association in that incorrect exemplars accompanied less accurate location memory. Alternatively, exemplar memory did not strongly modulate location memory for spatially consistent images, perhaps because participants could rely on prior knowledge to guide retrieval regardless of the fidelity of their memory of the specific image and its location.

Interestingly, exemplar memory also influenced how category typicality influenced bias, with stronger bias toward category neighbors for typical category members relative to atypical ones, but only for images with intact exemplar memory. Bias was not reliably affected by typicality for images with incorrect exemplar memory. Notably, we chose our sample size based on a power analysis that included estimates of exemplar memory accuracy, expecting that there would be fewer trials with incorrect exemplar memory and any comparisons of typicality by accuracy would be less powered when limited to trials with incorrect exemplar memory. Thus, we are confident that this analysis was sufficiently powered to identify differences by typicality for trials with incorrect exemplar memory. However, counter to our prediction, this difference in bias was not driven by changes in memory for the typical category members, as there was no indication of increased bias for typical category members with incorrect exemplar memory over ones with intact exemplar memory. We speculate that typical and atypical images with incorrect exemplar memory were similarly poorly encoded, with less processing of their category typicality, and thus retrieval of their locations was equally prone to influence by category neighbors. For stronger memories, as indicated by correct exemplar memory, retrieval of typical images was more biased than that of atypical images, suggesting a difference in processing that protects atypical category members from the influence of category neighbors. Indeed, while not significant, the retrieval of atypical category members was numerically less biased when their exemplar was correctly remembered relative to when it was forgotten.

Interestingly, how memory for details of the exemplars affects bias seems different from how reported confidence relates to bias, although both measures can be thought of as markers of memory strength. In this experiment as well as in Experiment 2, confidence in location memory did not interact with category typicality to influence bias (see Supplemental Results, Figure S3A–B). In contrast, images with intact exemplar memory exhibited a robust difference in bias by typicality, while images whose specific exemplar was forgotten did not. We discuss this potential discrepancy in more depth in the General Discussion.

In summary, we find that encoding in the context of a particular organizational structure alters memory for individual exemplars, in particular for details that are orthogonal or in opposition to the salient structure: memory for specific images and for random locations. Together, these results provide a possible explanation for why we observe differences in bias for location memory for typical and atypical category members, and opens up new testable questions about the relationship between the use of prior knowledge and the depth or efficiency of new encoding.

#### **Experiment 4**

In the prior three experiments, we found that location memory was more biased for typical category members relative to atypical ones. What characteristics of category typicality drive this effect? One possibility is that images of typical category members are more strongly influenced by other category members because they are more strongly associated (Collins & Loftus, 1975); for example, retrieval of typical words is more clustered than retrieval of atypical words in a free-recall task (Bousfield et al., 1958). Alternatively, typical category members may look more similar to other category members, and this increased visual similarity may make it easier to confuse the locations of typical category members with other images that happen to be near its cluster center. This explanation would not require prior semantic knowledge, but rather would apply to any stimuli that vary in visual similarity.

Experiment 4 aimed to adjudicate between these two possibilities. We developed a set of stimuli whose organization mimicked that of the categories employed in Experiments 1–3, but category typicality was determined by arbitrary assignment of colors to different manmade objects (Figure 5A). As an example, instead of using a category (e.g., birds) that comprised different basic level concepts (e.g., robins and eagles), in Experiment 4 a basic level category (e.g., lamps) comprised multiple exemplars of the concept (e.g., different styles of lamps), with an arbitrarily assigned color for the category (e.g., all lamps had a greenish hue). The color of each exemplar image was sampled from a distribution (e.g., each lamp could be more or less green). The exemplars assigned colors from the ends of this distribution were labeled "atypical" and the exemplars assigned colors from the center of this distribution were labeled "typical." Thus, "typical" exemplars looked more similar to other category members relative to "atypical" exemplars because they were closer in color. Note that this was the only sense in which the images were typical or atypical; the assignment of a particular exemplar to a specific color was randomly determined.

If typical images are retrieved closer to their similarly colored neighbors relative to atypical images, that would provide support for the notion that biases due to category typicality observed in Experiments 1–3 are driven in part by the visual

# Figure 5

Experiment 4



*Note.* (A) Typicality determined by color. The six vertical lines on the color bar indicate the center color of six categories. The distribution indicates the range of colors sampled for all exemplars of "lamp." Typical and atypical lamps reflect the distance between their color and the color at the center of the lamp distribution. (B) Average error by spatial consistency and group. (C) Average bias by color typicality and group. Gray lines indicate participants. Error bars indicate *SEM*.

\*\*\* p < .001. See the online article for the color version of this figure.

similarity of features across different category members. It would also provide support for the notion that biases in memory can be driven by newly learned, arbitrary mappings, like colors assigned to manmade objects, rather than stable semantic information, like knowledge of animal taxonomy and object function.

#### Method

# **Participants**

Seventy participants took part in Experiment 4, either in the experimental group (N = 35; 17 females; mean age = 34.6; SD = 8.1; range = 22–61 years) or the control group (N = 35; 12 females; mean age = 34.2; SD = 9.6; range = 25–64 years). The sample size was chosen for consistency with Experiments 1 and 2. Participants were recruited via AMT with the same procedures as in Experiments 1–3. Participants were paid \$5 upon completion of the experiment.

### **Materials**

Stimuli consisted of 70  $100 \times 100$  pixel images on white backgrounds (35 vehicles, 35 household objects; Brady et al., 2018). To mimic the six categories in Experiments 1–3, the images comprised exemplars of three vehicles (planes, trains, and cars) and three household objects (chairs, lamps, and clocks).

To mimic variance in category typicality, we created variance in how similar object exemplars were within each category by changing their colors (Figure 5A). First, we assigned each category a central color by choosing six equidistant points around a  $360^{\circ}$  color space (i.e., each value separated by  $60^{\circ}$ ). We then generated six normal distributions ( $\sigma = 20^{\circ}$ ) around these points. The width of these normal distributions ensured that the range of colors assigned to each category covered a largely unique color range. Next, we pseudorandomly sampled from each of these distributions to assign each image to a unique color that fell within its category's color range. Images with color values farthest from their category's central color were atypical category members, and images with colors closest to their category's central color were typical category members. To ensure that the range of colors assigned to spatially (in)consistent and (a)typical conditions was consistent across categories, the normal distribution for each category was divided into five bins, where the breaks between bins captured cumulative probabilities bounded by the following z-statistics: (inf, -1.405, -0.202, 0.202, 1.405, inf). Colors for atypical images were sampled from the two bins that comprised the tails of the distribution, colors for typical images were sampled from the bin comprising the center of the distribution, and colors for all spatially consistent images were evenly sampled from the two remaining bins. See Supplemental Figure 8 for an example of the spatial locations and image colors presented to a participant.

To ensure that any differences in category typicality were driven by color and not by other perceptual features of the images, the assignment of the six color ranges to the six categories and the assignment of colors to images within each category were randomized separately for six counterbalancing groups (see Supplemental Methods). In other words, the six groups viewed different combinations of image-color associations and different image exemplars associated with typical and atypical colors, minimizing the possibility that any observed influences of color typicality would be driven by idiosyncrasies of the other features of the images.

The number and locations of images in each category were exactly matched to the memory displays derived for Experiment 1. Specifically, locations that belonged to the bird category in Experiment 1 were assigned a color category with the same number of category members, and the locations of the (a)typical birds were now assigned (a)typically colored images. Thus, in addition to learning the locations of individual exemplars of an object, participants could learn that the object corresponded to a range of colors and a general location on the grid. Spatially "consistent" images were located near images of the same object in the same color range, and "inconsistent" images were in random locations. Critically, analogously to Experiments 1 and 2, the colors of half of the spatially inconsistent images were typical or sampled from the center of that category's color distribution, and the colors of the other half of the images were atypical, or sampled from the tails of the category's color distribution.

## Procedure

The timing, task, and instructions were identical to that of Experiment 1, with the identical confidence measure from Experiment 2.

#### Statistical Analyses

Analyses of error by spatial consistency and bias by category typicality are identical to those of Experiments 1–3. See the Supplemental Results for analyses of error (Figure S2C) and bias (Figure S3C) by confidence.

# Results

#### Spatial Consistency

We first examined whether the same differences in accuracy by spatial consistency observed in Experiments 1-3 would extend to stimuli organized by object type and color (Figure 5B). In line with those experiments, we predicted that memory would be more accurate for spatially consistent images relative to inconsistent ones (e.g., green lamps that were located near each other vs. ones that were in random locations). A 2 (group: experimental, control)  $\times$  2 (spatial consistency: consistent, inconsistent) ANOVA revealed a significant main effect of consistency, F(1, 68) = 76.87, p < .001,  $\eta_p^2 = .53$ , and no reliable main effect of group, F(1,68) = 3.68, p = .06,  $\eta_p^2 = .40$ . Critically, the ANOVA also revealed an interaction between group and consistency, F(1, 68) =16.52, p < .001,  $\eta_p^2 = 0.20$ . Like in Experiment 3, both the experimental group, t(34) = -7.78, p < .001, d = -1.32, and control group, t(34) = -4.16, p < .001, d = -0.70, exhibited greater error for spatially inconsistent images relative to consistent images, suggesting that the dense clustering of images in spatially consistent locations may have aided retrieval in both groups, but this difference in error by spatial consistency was greater for the experimental group relative to the control group. This suggests that the organization of images by object and color further aided memory for images whose locations were consistent with that spatial organization.

#### Color ("Category") Typicality

As a reminder, our critical prediction was that retrieval of typically colored images would be more biased toward the similarly colored exemplars of the same object relative to atypically colored images. This finding would be analogous to the biases due to category membership seen in the first three experiments. However, there was no reliable difference in bias for typical images relative to atypical images in the experimental group, t(34) = 1.46, p = .15, d = 0.25, or in control group, t(34) = 0.76, p = .45, d =.13 (Figure 5C). Furthermore, a 2 (group: experimental, control)  $\times$ 2 (typicality: typical color, atypical color) ANOVA did not reveal a main effect of typicality, F(1, 68) = 2.38, p = .13,  $\eta_p^2 = .03$ , or a significant typicality by group interaction, F(1, 68) = 0.17, p =.68,  $\eta_p^2 = .002$ . Interestingly, this analysis revealed a main effect of group, F(1, 68) = 18.75, p < .001,  $\eta_p^2 = 0.51$ , with more bias regardless of category typicality in the experimental group relative to the control group, t(68) = 4.33, p < .001, d = 1.04. Taken together with the error analyses, this suggests that when images were organized by color and object type, the retrieval of spatially inconsistent images was biased toward the center of their object clusters, even though their accuracy did not differ from subjects whose images were not organized by object type and color.

As typically colored and atypically colored images did not differ in the proportion of error that was biased toward their category cluster, it may seem like the color manipulation simply was not effective, and that participants did not meaningfully encode the colors of each image or use color information during retrieval. However, we uncovered one piece of evidence that this manipulation may have had a limited effect. When considering an image's unadjusted bias (Figure 1D), we found that typically colored images were retrieved closer to their category center relative to atypically colored images in the experimental group, t(34) = 2.05, p = .048, d = .35, and not in the control group, t(34) = -0.30, p = .76, d = -.05. However, there was no reliable interaction between group and typicality,  $F(1, 68) = 1.91, p = .17, \eta_p^2 = 0.03$ . While not robust, this provides some evidence that participants attended to the colors of each image during encoding and used this information to a limited degree during retrieval.

Finally, for comparison to Experiments 1–3, we also examined location memory accuracy among the spatially inconsistent images to understand how the typicality of an image's color, relative to the mean color of its object category, influenced location memory (Figure S1). As a reminder, retrieval of typical category members was less accurate than retrieval of atypical category members in the prior experiments. We asked whether there were analogous differences in error by the typicality of the colors assigned to each image. A 2 (group: experimental, control)  $\times$  2 (typicality: typical color, atypical color) ANOVA revealed no reliable main effects or interaction between typicality and group, all *F* < 0.49, all *p* > .48. The lack of an effect suggests that memory accuracy was not influenced by the typicality of an image's color.

#### Discussion

The aim of Experiment 4 was to understand whether the biases in memory by category typicality observed in Experiments 1–3 could be driven by visual similarity of the images. Over the course of encoding specific image-location associations, participants also learned that exemplars of a certain type of object, like lamps, tended to be located in certain areas of the screen and associated with a certain color. This organization influenced the retrieval of specific images, such that images in random locations were retrieved less accurately. Note that differences in accuracy here could be driven either by images clustering by object type, by color, or by a combination of the two.

The critical comparison was of bias in retrieval based on whether the arbitrary colors of encoded objects were more or less similar to the mean color of each object set. This continuous variation in color within an object set mirrors the graded manner in which category members vary in their similarity to a central tendency or prototype of each category. Analogously to the typical and atypical category members in spatially inconsistent locations in Experiments 1-3, images whose colors were closest or farthest from the mean color of that object were randomly located. We found weak evidence that typically colored images were retrieved closer to their category clusters, but only when employing a different bias measure that did not account for error. When employing the identical measures used in Experiments 1-3, color typicality did not influence accuracy or bias by their category clusters. This suggests that visual similarity alone cannot explain the bias effects observed in Experiments 1-3. Rather, bias in retrieval in Experiments 1-3 may be driven primarily by the strength of associations between category members, but also in small part by their visual similarity.

One reason why visual similarity did not robustly influence memory may be that the task was more difficult than in Experiments 1-3, because participants needed to encode the locations of many exemplars of the same object sets, rather locations of unique animals and objects. Differences in difficulty could also explain the strong main effect of group on bias in this experiment, where the experimental group exhibited more bias relative to the control group regardless of category typicality. However, accuracy in Experiment 4 was not reliably worse than in Experiment 1 when averaging across all conditions (see Comparisons Across Experiments section), which suggests that Experiment 4 was not meaningfully more difficult. Another possibility is that the feature that gave structure to each category was based purely on color, which was newly learned and arbitrarily assigned. In contrast, images in Experiments 1-3 were organized by multiple features of animals and objects that likely were learned long before participants began the experiment. Moreover, features that give rise to visual similarity may also be correlated with other dimensions of similarity. For example, a cardinal and a blue jay look similar because they both have wings, a perceptual feature that also serves the same function. Thus, differences in images' typicality by color may have been less well encoded and thus did not bias retrieval as strongly as in the prior experiments.

#### **Comparisons Across Experiments**

The four experiments varied in the number of images and categories and the manner in which their locations were organized (see Table 1). In the following section, we provide a direct comparison of results across the four experiments.

#### **Error by Spatial Consistency**

First, we investigated whether average error differed across the four experiments as a function of spatial consistency. We implemented a mixed-effects model with spatial consistency (consistent, inconsistent), group (experimental, control), experiment (Experiments 1-4), and their interactions as discrete independent vari-

Table 1	
Similarities and Differences Across the Four E	xperiments

Design	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Images	70 animals and objects	160 animals and objects	70 animals and objects with similar lures	70 exemplars of 6 objects
Categories	3 animal and 3 object categories, derived from odd-man-out	4 animal and 4 object categories, predetermined	Same as Experiment 1	3 vehicles and 3 household objects, arbitrary colors
Typicality	Derived from odd-man-out	Derived from list ranking task	Same as Experiment 1	Distance between image color and the center color of its object set
Locations	Derived from odd-man-out	Clustered by categories, locations evenly distributed	Same as Experiment 1	Same as Experiment 1
Retrieval	Drag image to encoded location	Same as Experiment 1	Choose encoded image, drag to location	Same as Experiment 1
Confidence	None	Rate confidence of location	Rate confidence of image Rate confidence of location	Rate confidence of location

ables and average error as the dependent variable. This model revealed no main effect of experiment, F(3, 440) = 1.24, p = .29; no main effect of group, F(1, 440) = 1.72, p = .19; and no group by experiment interaction, F(3, 440) = 0.72, p = .54. This suggests that the average memory accuracy across all images was not reliably different across the four experiments or across the experimental and control groups, despite many differences across the experimental procedures.

This same model was used to investigate whether the relationship between accuracy and spatial consistency differed across the four experiments. Along with a main effect of consistency, F(3,(440) = 197.85, p < .001, the model revealed a strong interaction between consistency and group, F(1, 440) = 82.36, p < .001, reflecting the greater difference in error by spatial consistency for the experimental group relative to the control group that was observed in all four experiments. Critically, there was no threeway interaction, F(3, 440) = 1.14, p = .33, suggesting that the extent to which memory accuracy was more influenced by spatial consistency in the experimental groups over the control groups was not meaningfully different across the four experiments. There was also an interaction between consistency and experiment, F(3,(440) = 6.38, p < .001, which reflected a larger difference in error by consistency in Experiment 4 relative to Experiments 1-3, when collapsing across experimental and control groups. This is because error in the control group was also modulated by spatial consistency, but to a lesser extent than in the experimental group. As a reminder, in Experiments 1-3, participants learned that most members of a category tended to be clustered in the same general location, while in Experiment 4, most exemplars of the same object within the same color range were clustered. As participants in Experiment 4 needed to discriminate and encode many exemplars of an object, it may be that participants' overall memory in this experiment was driven less by object and color information and more by the uneven density of spatial locations present in both the experimental and control groups. See Figure 6A for a comparison of standardized effect sizes for each group and experiment.

# **Bias By Typicality**

We next asked whether there were differences in bias by category typicality across the four experiments. To do this, we computed a mixed-effects model with category typicality (typical, atypical), group (experimental, control), experiment (Experiments 1-4), and their interactions as discrete fixed effects and average bias as the dependent variable. This analysis revealed main effects of typicality, F(1, 440) = 11.73, p < .001, and group, F(1, 440) =19.75, p < .001, which were qualified by the expected interaction between typicality and group, F(1, 440) = 15.83, p < .001. This interaction reflects greater bias for typical category members over atypical ones across all experimental groups. Critically, there was no three-way interaction, F(3, 440) = 1.38, p = .25. This suggests that despite the numerous differences across the four experiments, the extent to which category typicality differentially biased memory in the experimental groups over the control groups did not meaningfully differ. There was also a significant main effect of experiment, F(3, 440) = 7.26, p < .001, qualified by a group by experiment interaction, F(3, 440) = 5.06, p = .002. This interaction reflected the finding that when collapsing across typicality, experimental groups exhibited more bias than did control groups only in Experiments 2 and 4. See Figure 6B for a comparison of effect sizes of bias by typicality for both the experimental and control groups in each experiment.

#### **General Discussion**

Across four experiments, we aimed to understand how category typicality governs distortions in new episodic memories when prior knowledge about categories is used to scaffold new learning. To do this, we developed a novel location memory task, where participants encoded associations between images and locations. Critically, most images were spatially clustered by their category membership, such that participants could learn associations between an image's category and its spatial location, but some images were in random locations. We predicted that retrieval of episodic memories would be supported by a mixture of recalled details of the original event as well as the new knowledge of the general location of each category, which could only be used if participants recruited their semantic knowledge of animal and object categories as they encoded specific image-location associations. Indeed, we found that this new information influenced both the accuracy and distortion of new memories. First, memory was more accurate for images located near others from the same

# Figure 6

Effect Sizes Across Experiments



*Note.* (A) Standardized difference in mean error for spatially consistent images versus spatially inconsistent images. (B) Standardized difference in mean bias for typical versus atypical images. Points represent mean effect size and lines represent bootstrapped 95% confidence intervals. Zero on the *x*-axis reflects no difference in error by consistency or bias by typicality. See the online article for the color version of this figure.

category. Second, for images that were in random locations, participants made systematic errors: Typical category members were retrieved as closer to neighbors of the same category than atypical category members. These results replicated across multiple experiments despite differences in the number and type of categories presented, method of determining typicality, specific images used, and mapping between category membership and spatial location. Below, we discuss these findings through the lens of the category adjustment model (CAM), position our observations in research on category influences on encoding and retrieval, and consider some limitations of the current design and opportunities for future work.

# Evidence For and Beyond the Category Adjustment Model

We first focus on the finding that memory for spatially consistent images was more accurate than memory for images in random locations, and that this difference was further modulated by reported confidence. These findings provide strong evidence for the theory that episodic retrieval is a reconstructive process (Bartlett, 1932; Schacter et al., 1998). Formalized as the CAM, this framework posits that if a newly experienced event is consistent with prior knowledge, combining the details of that event with prior knowledge is thought to enhance its retrieval by reducing noise (Huttenlocher et al., 2000). In our experiments, an image located near category neighbors could be retrieved by combining memory for its location with new knowledge of the locations of images from the same category, resulting in more accurate retrieval of these images' associated locations. Our findings extend beyond a large and diverse body of work showing that prior knowledge facilitates new memories (Bransford & Johnson, 1972; Posner & Keele, 1968; Schulman, 1974) by demonstrating that memory accuracy can also be influenced by newly acquired knowledge (e.g., a category's spatial location). The memory reconstruction framework also considers the relative strengths of prior knowledge and memory for specific events, such that retrieval of an event will rely more on prior knowledge with greater uncertainty of the event memory (Brady et al., 2018). In Experiments 2 and 3, we found an interaction between confidence and spatial consistency, such that the effect of new knowledge on accuracy was weaker for images whose locations were confidently remembered. Furthermore, confident memories were less biased toward category members. Both observations provide support for CAM.

However, our findings related to category typicality diverge from CAM's predictions. In these cases, integration of eventspecific and category information would result in a distorted memory that is somewhere between the veridical event and what would have been expected from prior knowledge. Notably, the vast majority of demonstrations of such biases find that there is greater bias for atypical items over typical ones (Hemmer & Steyvers, 2009; Huttenlocher et al., 1991; Persaud & Hemmer, 2016). In our experiments, we found a seemingly opposite effect-memory for the locations of typical category members was more biased toward the general location of its category. This finding is more in line with what would be predicted by spreading activation theory, in that typical category members are more biased toward the locations of other category members because of their stronger associations with their category. However, we argue that our findings are not discrepant with CAM but instead add nuance to the model by exposing a limitation of how it is often tested. In most empirical experiments providing evidence for CAM, the category information that influences individual estimations is organized along the same dimension along which memory is tested. In other words, an item's tested distance from its category center is perfectly correlated with its similarity to the category prototype. As an example, when participants encoded line segments of various lengths, estimates of the length of each segment were biased toward the mean length of segments already viewed. Thus, the length of an atypically long segment is both furthest from its prototype and least similar to it (e.g., Allred et al., 2016; Duffy et al., 2010). While this is a valid assumption when considering unidimensional stimuli or stimuli with correlated dimensions (Crawford et al., 2006; Dubova & Goldstone, 2020; Huttenlocher & Hedges, 1994), it does not consider memories that vary along conflicting dimensions, like a typical category member (short distance to prototype) encoded far from its category's newly learned spatial location. In the current experiments, both typical and atypical category members were located in random locations, far from the center of their category's location. Thus, the similarity to a category prototype was uncoupled from its physical distance. Moreover, unlike past experiments, the new category-location mapping that biased memory was not intrinsically related to the images (e.g., nothing about the concept of a "spatula" implies that it should be located on the top right corner of a grid). However, by associating these locations with the category of the images, participants learned to treat location as a new "feature" of the image associated with its category membership. Thus, prior knowledge can help to organize the acquisition of new information, but this new information gives rise to larger biases for typical category members.

# Shaping Typical and Atypical Memories in the Context of New Knowledge

One important avenue for future work is to understand when this reconstruction process supports memory. CAM posits that bias emerges during retrieval, when integrating the noisy, but unbiased, details of the experience with other prior knowledge (Hemmer & Persaud, 2014; Huttenlocher et al., 2000). At the same time, however, there is also evidence that prior knowledge shapes encoding, as it can enhance reading comprehension and various perceptual tasks (Bransford & Johnson, 1972; Posner et al., 1967; Ratcliff & McKoon, 1995). Furthermore, drawing attention to category information can magnify differences in dimensions that explain category membership (Goldstone, 1994, 1995; Levin & Beale, 2000; Livingston et al., 1998). As such effects are evident even during perception, they likely do not rely on retrieval computations. Similarly, neuroscientific evidence suggests a role for prior knowledge both when encoding (Bein et al., 2014; Tse et al., 2011; van Kesteren, Fernández, et al., 2010), and retrieving (van Kesteren, Rijpkema, et al., 2010) new memories. Taken together, this suggests that prior knowledge likely shapes new memories both at encoding and retrieval.

In Experiment 3, we tested one way that memory may be influenced by integrating prior category knowledge with new information, which is through a change in the depth of encoding for typical and atypical category members (Craik & Lockhart, 1972; Craik & Tulving, 1975). In the current experiments, recognizing the utility of category information when encoding locations

may have encouraged a more relational encoding strategy, which is thought to bring attention to atypical or distinctive items (Dunlosky et al., 2000; Hunt & Lamb, 2001). For example, when studied as part of a salient schema or category structure, inconsistent items tend to receive more study time (G. R. Loftus & Mackworth, 1978; Stern et al., 1984), changes to inconsistent stimuli are more likely to be noticed (Friedman, 1979; Goodman, 1980; Heider, 1946; Schank & Abelson, 1977; Sentis & Burnstein, 1979), and inconsistent items are more often remembered (e.g., Bower et al., 1979; Sakamoto & Love, 2004). On the other hand, the presence of category information may result in more efficient encoding of items that are not distinctive in the context of the salient organizational scheme, but this may impair in-depth processing of their idiosyncratic details (Sweegers et al., 2015; Sweegers & Talamini, 2014). To test this prediction, we added an exemplar memory test to assess the depth of encoding of typical and atypical category members. Performance on this test depended on the encoding context: When category information predicted the locations of most images, memory for atypical category members was greater than memory for typical ones, but when category information was not relevant during encoding (in the control group), there was no difference in exemplar memory by typicality. Furthermore, we found that error in location memory for typical category members was higher in Experiments 1-3 (Figure S1). Better memory for both the locations and specific exemplars of atypical images suggests that category typicality influences the depth of their encoding, where typical category members are processed in a way that discounts their idiosyncratic features, while the distinctiveness of atypical items relative to other category neighbors draws processing to their specific details.

Finally, a novel feature of our design was our ability to test whether separate elements of the same event can dictate bias by new learning. In our experiments, each trial was composed of multiple "episodic" elements, including the visual features of an image and its spatial location, and "semantic" elements, including the image's category membership and typicality. Because we were able to track memory for the episodic elements, we found that memory for an image's exemplar seemed to be unrelated to memory for its location. We reported that, in Experiments 2 and 3, less confident responses were more biased toward category centers, and typical category members were more biased than atypical ones, but confidence in location memory did not interact with category typicality (Figure S3A-B). In contrast, typicality interacted with exemplar memory in that images with intact exemplar memory exhibited a robust difference in bias by typicality, while images whose specific exemplar was forgotten did not. At first glance, these results seem discrepant, as exemplar memory and location confidence both seem to index the strength of memory for specific events, suggesting that both low-confident responses and incorrect exemplar memory should give rise to more bias due to a stronger influence of new knowledge of the category locations. However, our results suggest that memory for different elements of an event may differ in their susceptibility to this bias. Memory for the specific exemplar may actually be a better index of the semantic element of each memory, where weaker memory disrupts the biasing effect of other category members, while location confidence corresponds to an episodic element, which is more prone to distortion by category members if it is weak. Analogously to prior work showing that episodic memory can be affected by multiple

sources of prior knowledge (Hemmer & Steyvers, 2009), it may be that memory of an event is composed of multiple, separable sources of event-specific information that differently interact with newly acquired information.

#### **Reconciliations With the Category Adjustment Model**

The basic framework from which CAM is derived posits that biases in retrieval are explained by the Bayesian integration of category knowledge with fine-grained details of a specific event. In this section, we discuss how this model could be extended and reconfigured to account for our findings.

One approach would involve keeping the estimate of category locations consistent across all trials and adjusting the uncertainty of estimates for specific locations, with less uncertainty for atypical category members. This approach is inspired by the findings of Experiment 3, showing better exemplar memory for atypical category members over typical ones. If deeper encoding of atypical category members leads to less uncertainty in the estimates their locations, this would result in more biases for typical category members, consistent with our observations.

An alternative way that CAM could be modified is by adjusting the uncertainty corresponding to participants' estimates of the newly learned category locations. To account for typicality effects, this estimate could be assigned greater uncertainty (wider distribution) when retrieving the locations of atypical items and less uncertainty (narrower distribution) when estimating the locations of typical ones. This scaling would represent the different extent to which participants would expect a typical versus atypical category member to be located near other category members. Thus, integrating this estimate with the estimate of the locations of each image would result in more bias for typical category members over atypical ones. Developing characterizations of participants' knowledge of the category locations, perhaps by asking them to place novel typical and atypical category members onto the grid, could be useful for testing this model.

#### Limitations and Avenues for Future Work

Our priority was to establish whether any aspect of semantic memory that could be mapped to new knowledge would influence distortions in episodic memory. To achieve this aim, we prioritized developing a distinction between category members that was robust across participants, choosing to leverage typicality within categories for the reasons described in the Introduction. However, typicality is one of many ways that category membership is organized and measured. The graded structure of categories is also well characterized by the frequency that its members are encountered in the environment; their central tendency, or similarity to other category members (or, inversely, their distinctiveness from all other category members); their accessibility, or output dominance, characterized as the frequency with which an item is produced as an example of a category. These characteristics are often highly correlated, such that typical category members are more easily accessible (Schmidt, 1996). Indeed, past observations of clustering in free recall are well explained by category typicality, but more variance is explained by accessibility (Smith et al., 2000). At the same time, increased false alarms for typical items in studied word lists (Roediger & McDermott, 1995; Smith et al., 2000) could occur because typical category members more easily come to mind when cued by a category, as reflected by their greater output dominance, or production frequency (Barsalou, 1985; DeSoto & Roediger, 2014; Keller & Kellas, 1978). Because we did not control for familiarity or accessibility when choosing our stimuli, it is possible that these dimensions of semantic knowledge could be used to obtain similar biases in episodic memory. Either way, our results demonstrate that the internal structure of a category can explain the extent of distortions in new episodic memories. One promising future avenue of research may to understand potential commonalities and distinctions in how semantic knowledge distorts new memories along these multiple dimensions.

We also asked whether perceptual confusability could explain our observed results. Because typical category members by definition share more features with others (Rosch & Mervis, 1975), they may be more confusable if the shared features are visual, such that locations of typical category members may have been more likely to be swapped for the location of a category neighbor that happened to be near its cluster center. This explanation could be applied to any set of stimuli that vary in visual similarity. We offer two explanations why this may not account for our effects. First, bias driven by visual confusability leads to the prediction that bias would be greater for animals over objects, as visual similarity is less of an organizing dimension for manmade artifacts. However, we find no systematic differences in bias by category typicality in animal or object categories (see Supplemental Results for more details). Second, we designed Experiment 4 to quantify bias solely as a function of visual similarity. To do this, we assigned exemplars of different manmade objects with colors that determined their typicality. We found that, relative to atypically colored images, typically colored images were retrieved closer to their respective cluster, but only when using a measure of unadjusted bias, unlike what was used in Experiments 1-3. It is worth noting that effects observed in Experiment 4 may be weaker because the object-color associations were new and arbitrary, and thus may be more difficult to retain relative to the well-known features that determined typicality in Experiments 1-3. Regardless, this result suggests that the observed differences in bias by category typicality is driven in some small part by differences in visual similarity, but visual similarity is likely one of many features that differentiates typical and atypical category members and gives rise to differences in their bias.

An important feature of our design is that retrieval could be quantified as a continuous variable, allowing us to quantify both the direction and magnitude of error for each trial. However, this measurement is unable to disentangle the types of error in memory that underpin the observed differences in bias by category typicality; for instance, bias in memory may be characterized as either a relatively small distortion, where error is minimal but systematically influenced by category knowledge, or as a swap error, where bias toward an image's category center is due to conflating memory for its location with the location of another image near the cluster center. It seems likely that location memory for typical category members would be more likely to be swapped in such a manner, thus explaining the increased bias toward the category center, but because the locations of the images necessarily were densely distributed, it is not possible to tease apart whether swap errors occurred (cf. Lew et al., 2016). Although we did not originally design the reported experiments to adjudicate between these possibilities, we attempted to characterize participants' retrieval in a manner that captured both types of error by binning each retrieved location as a function of the relative distance between its encoded location and cluster center (Figure S5). We found that across all experiments in which stimuli were organized by category, retrieval was closer to their encoded locations than to their respective cluster centers. Although typical category members were overall more likely to be retrieved closer to their category center relative to atypical ones, they were still most likely to be retrieved closer to their actual encoded locations. While this exploratory analysis should be interpreted with caution, it hints at the possibility that the difference in bias by category typicality was driven by small differences in the magnitude of error, rather than qualitative differences in how these category members were retrieved.

#### Conclusions

We have presented an investigation of how prior semantic knowledge can scaffold new learning, leading to distortions in new episodic memories. This work demonstrates that semantic knowledge and episodic memory are closely intertwined in supporting new learning and offers an opportunity to better understand how information supported by these two systems is integrated in the context of new memory formation.

#### Context

The first author has previously studied systems-level memory consolidation mechanisms (Murty et al., 2017; Tompary et al., 2015) and in particular how consolidation allows for integration across overlapping events (Tompary & Davachi, 2017). The last author has a longstanding interest in the organization of concepts (Musz & Thompson-Schill, 2015; Yee & Thompson-Schill, 2016) as well as mechanisms underlying new learning (e.g., Coutanche et al., 2013; Karuza et al., 2016; Leshinskaya & Thompson-Schill, 2020). The current experiments were conceived from our shared interest in understanding how prior semantic knowledge influences the acquisition of new information, and in particular how memory for specific events may be distorted by relying on prior knowledge to support new learning. The predictions generated for these experiments bridged research on memory reconstruction with theories explaining category typicality effects across a broad array of cognitive tasks. In these experiments, we show that when prior knowledge is recognized as a useful dimension along which to learn new information, memories of specific events are systematically distorted according to their organization in that dimension.

#### Data and Code Availability

Stimuli, raw data, and analysis code that support the findings of the experiments are available at https://osf.io/8ugek. Raw data and analysis code dedicated to stimulus development will be made available upon request.

# References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, 93(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

- Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review*, 23(6), 1825–1831. https://doi.org/10.3758/s13423-016-1039-0
- Bae, G-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144(4), 744–763. https://doi.org/10.1037/ xge0000076
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629
- Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge University Press.
- Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 80(3, Pt. 2), 1–46. https:// doi.org/10.1037/h0027577
- Bein, O., Reggev, N., & Maril, A. (2014). Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia*, 64, 320–330. https://doi.org/10.1016/j .neuropsychologia.2014.09.046
- Bejjani, C., & Egner, T. (2019). Spontaneous task structure formation results in a cost to incidental memory of task stimuli. *Frontiers in Psychology*, 10, 2833. https://doi.org/10.3389/fpsyg.2019.02833
- Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. (1958). Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychological Reports*, 4(1), 39–44. https://doi.org/10 .2466/pr0.1958.4.g.39
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11(2), 177–220. https://doi.org/10.1016/ 0010-0285(79)90009-4
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384–392. https://doi.org/10.1177/ 0956797610397956
- Brady, T. F., Schacter, D. L., & Alvarez, G. A. (2018). The adaptive nature of false memories is revealed by gist- based distortion of true memories. *PsyArXiv*. https://doi.org/10.31234/osf.io/zeg95
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717–726. https:// doi.org/10.1016/S0022-5371(72)80006-9
- Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 118–160). Lawrence Erlbaum Associates Publishers.
- Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13(2), 207–230. https://doi.org/10.1016/ 0010-0285(81)90008-6
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5(5), e10773. https://doi.org/10.1371/journal.pone.0010773
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Per-spectives on Psychological Science*, 6(1), 3–5. https://doi.org/10.1177/ 1745691610393980
- Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. https://doi.org/10 .1037/0033-295X.82.6.407

- Collins, A., & Quillian, R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8(2), 240–247. https:// doi.org/10.1016/S0022-5371(69)80069-1
- Connolly, A. C., Gleitman, L. R., & Thompson-Schill, S. L. (2007). Effect of congenital blindness on the semantic representation of some everyday concepts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), 8241–8246. https://doi.org/10.1073/ pnas.0702812104
- Coutanche, M. N., Gianessi, C. A., Chanales, A. J. H., Willison, K. W., & Thompson-Schill, S. L. (2013). The role of sleep in forming a memory representation of a two-dimensional space. *Hippocampus*, 23(12), 1189–1197. https://doi.org/10.1002/hipo.22157
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72) 80001-X
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268
- Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin & Review*, 13(2), 245–250. https://doi.org/10.3758/BF0 3193838
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, 8(3), e57410. https://doi.org/10.1371/journal.pone .0057410
- DeSoto, K. A., & Roediger, H. L., III. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science*, 25(3), 781–788. https://doi.org/10.1177/0956797613516149
- Deyne, S. D., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, 40(4), 1030–1048. https://doi.org/10.3758/BRM .40.4.1030
- Djalal, F. M., Ameel, E., & Storms, G. (2016). The typicality ranking task: A new method to derive typicality judgments from children. *PLoS ONE*, *11*(6), e0157936. https://doi.org/10.1371/journal.pone.0157936
- Dubova, M., & Goldstone, R. (2020). The influences of category learning on perceptual reconstructions. *PsyArXiv*. https://doi.org/10.31234/osf.io/ 93w4j
- 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. https:// doi.org/10.3758/PBR.17.2.224
- Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychol*ogy: Learning, Memory, and Cognition, 26(3), 649–657. https://doi.org/ 10.1037/0278-7393.26.3.649
- Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus typicality determines how broadly fear is generalized. *Psychological Science*, 25(9), 1816– 1821. https://doi.org/10.1177/0956797614535401
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychol*ogy: General, 108(3), 316–355. https://doi.org/10.1037/0096-3445.108 .3.316
- Goldstone, R. L. (1994). 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. (1995). Effects of categorization on color perception. *Psychological Science*, 6(5), 298–304. https://doi.org/10.1111/j.1467-9280.1995.tb00514.x

- Goodman, G. S. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology*, 12(4), 473–495. https://doi.org/10 .1016/0010-0285(80)90017-1
- Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. *The Quantitative Methods for Psychology*, 14(4), 242–265. https://doi.org/10 .20982/tqmp.14.4.p242
- Greenberg, M. S., & Bjorklund, D. F. (1981). Category typicality in free recall: Effects of feature overlap or differential category encoding? *Journal of Experimental Psychology: Human Learning and Memory*, 7(2), 145–147. https://doi.org/10.1037/0278-7393.7.2.145
- Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, 21, 107–112. https://doi.org/10.1080/00223980.1946 .9917275
- Hemmer, P., & Persaud, K. (2014). Interaction between categorical knowledge and episodic memory across domains. *Frontiers in Psychology*, 5, 584. https://doi.org/10.3389/fpsyg.2014.00584
- Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin* & *Review*, 16(1), 80–87. https://doi.org/10.3758/PBR.16.1.80
- Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461–469. https://doi.org/10.2307/2012819
- Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(6), 1359–1366. https://doi.org/10.1037/0278-7393.27.6.1359
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, 32(4), 421–445. https://doi.org/10.1006/jmla.1993.1023
- Huttenlocher, J., & Hedges, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review*, 101(1), 157–165. https://doi.org/10.1037/0033-295X.101.1.157
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, 98(3), 352–376. https://doi.org/10.1037/0033-295X.98.3 .352
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: Gen*eral, 129(2), 220–241. https://doi.org/10.1037/0096-3445.129.2.220
- Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: Influences of network topology on human learning. *Trends in Cognitive Sciences*, 20(8), 629–640. https:// doi.org/10.1016/j.tics.2016.06.003
- Keller, D., & Kellas, G. (1978). Typicality as a dimension of encoding. Journal of Experimental Psychology: Human Learning and Memory, 4(1), 78-85. https://doi.org/10.1037/0278-7393.4.1.78
- Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, 45(4), 905–927. https://doi.org/10.3758/s13428-013-0330-5
- Kirwan, C. B., Jones, C. K., Miller, M. I., & Stark, C. E. L. (2007). High-resolution fMRI investigation of the medial temporal lobe. *Human Brain Mapping*, 28(10), 959–966. https://doi.org/10.1002/hbm.20331
- Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience*, 33(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). ImerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(1), 1–26. https://doi.org/10.18637/jss.v082.i13
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863
- Leeper, T. J. (2017). MTurkR: Access to Amazon Mechanical Turk Requester API via R (R package version 0.8.0.) [Computer software]. https://github.com/cloudyr/MTurkR

- Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means [Computer software]. https://CRAN.R-project.org/package= emmeans
- Leshinskaya, A., & Thompson-Schill, S. L. (2020). Transformation of event representations along middle temporal gyrus. *Cerebral Cortex*, 30(5), 3148–3166. https://doi.org/10.1093/cercor/bhz300
- Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*, 62(2), 386–401. https://doi.org/10.3758/BF03205558
- Lew, T. F., Pashler, H. E., & Vul, E. (2016). Fragile associations coexist with robust memories for precise details in long-term memory. *Journal* of Experimental Psychology: Learning, Memory, and Cognition, 42(3), 379–393. https://doi.org/10.1037/xlm0000178
- 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
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4(4), 565–572. https:// doi.org/10.1037/0096-1523.4.4.565
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309
- 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
- McCloskey, M. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 485–502. https://doi.org/10.1016/S0022-5371(80)90330-8
- Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, 7(3), 283–284. https://doi.org/10.3758/ BF03337190
- Morris, P. E., & Fritz, C. O. (2013). Effect sizes in memory research. *Memory*, 21(7), 832–842. https://doi.org/10.1080/09658211.2013 .763984
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 70–84. https://doi.org/10.1037/0278-7393.11.1.70
- Murty, V. P., Tompary, A., Adcock, R. A., & Davachi, L. (2017). Selectivity in postencoding connectivity with high-level visual cortex is associated with reward-motivated memory. *The Journal of Neuroscience*, 37(3), 537–545. https://doi.org/10.1523/JNEUROSCI.4032-15 .2016
- Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*, 76, 41–51. https://doi.org/10.1016/j.neuropsychologia.2014.11.029
- Nosofsky, R. M. (1986). Attention, similarity, and the identification– categorization relationship. *Journal of Experimental Psychology: Gen*eral, 115(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology*, 17(4), 273–281. https://doi.org/10.1080/17470216508416445
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97(2), 185–200. https://doi.org/10.1037/0033-295X.97.2.185

- Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, 88, 1–21. https:// doi.org/10.1016/j.cogpsych.2016.05.003
- Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (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. https://doi.org/10 .1037/h0025953
- Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, 86(4), 777–803. https://doi.org/10.1037/0033-2909.86.4.777
- Ratcliff, R., & McKoon, G. (1995). Bias in the priming of object decisions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(3), 754–767. https://doi.org/10.1037/0278-7393.21.3.754
- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.Rproject.org/
- Rips, L. J. (1975). Inductive judgments about natural categories. Journal of Verbal Learning and Verbal Behavior, 14(6), 665–681. https://doi.org/ 10.1016/S0022-5371(75)80055-7
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20. https://doi.org/10.1016/S0022-5371(73) 80056-8
- 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(4), 803–814. https:// doi.org/10.1037/0278-7393.21.4.803
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9
- Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 491–502. https://doi.org/10.1037/0096-1523.2.4.491
- Russo, N., Hagmann, C. E., Andrews, R., Black, C., Silberman, M., & Shea, N. (2018). Validation of the C.A.R.E. stimulus set of 640 animal pictures: Name agreement and quality ratings. *PLoS ONE*, *13*(2), e0192906. https://doi.org/10.1371/journal.pone.0192906
- Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, 133(4), 534–553. https://doi.org/10.1037/0096-3445.133.4.534
- Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, 49(1), 289–318. https://doi.org/10.1146/annurev.psych.49.1.289
- Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals, and understanding. Erlbaum.
- Schmidt, S. R. (1996). Category typicality effects in episodic memory: Testing models of distinctiveness. *Memory & Cognition*, 24(5), 595– 607. https://doi.org/10.3758/BF03201086
- Schulman, A. I. (1974). Memory for words recently classified. Memory & Cognition, 2(1), 47–52. https://doi.org/10.3758/BF03197491
- Sentis, K. P., & Burnstein, E. (1979). Remembering schema-consistent information: Effects of a balance schema on recognition memory. *Jour*nal of Personality and Social Psychology, 37(12), 2200–2211. https:// doi.org/10.1037/0022-3514.37.12.2200
- Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition*, 28(3), 386–395. https://doi.org/10.3758/BF03198554
- Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*,

51(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12 .014

- Stern, L. D., Marrs, S., Millar, M. G., & Cole, E. (1984). Processing time and the recall of inconsistent and consistent behaviors of individuals and groups. *Journal of Personality and Social Psychology*, 47(2), 253–262. https://doi.org/10.1037/0022-3514.47.2.253
- Sweegers, C. C. G., Coleman, G. A., van Poppel, E. M., Cox, R., & Talamini, L. M. (2015). Mental schemas hamper memory storage of goal-irrelevant information. *Frontiers in Human Neuroscience*, 9, 629. https://doi.org/10.3389/fnhum.2015.00629
- Sweegers, C. C. G., & Talamini, L. M. (2014). Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex*, 59, 49–61. https://doi.org/10.1016/j.cortex.2014.07.006
- Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*, 96(1), 228–241.e5. https://doi.org/10.1016/j .neuron.2017.09.005
- Tompary, A., Duncan, K., & Davachi, L. (2015). Consolidation of associative and item memory is related to post-encoding functional connectivity between the ventral tegmental area and different medial temporal lobe subregions during an unrelated task. *The Journal of Neuroscience*, 35(19), 7326–7331. https://doi.org/10.1523/JNEUROSCI.4816-14.2015
- Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., & Morris, R. G. M. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science*, 333(6044), 891–895. https://doi.org/10.1126/science.1205274
- Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation*, 12(6), 587–595. https://doi.org/10.3758/BF03201848

- van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernandez, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience*, 30(47), 15888–15894. https://doi.org/10 .1523/JNEUROSCI.2674-10.2010
- van Kesteren, M. T. R., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), 7550–7555. https://doi.org/10.1073/pnas.091 4892107
- von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung*, 18(1), 299–342. https://doi.org/10.1007/BF02409636
- Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, 63(6), 410–424. https://doi.org/10.1037/h0022001
- Whitney, P., Cocklin, T. G., Juola, J. F., & Kellas, G. (1983). A reassessment of typicality effects in free recall. *Bulletin of the Psychonomic Society*, 21(4), 321–323. https://doi.org/10.3758/BF03334722
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag. https://ggplot2.tidyverse.org
- Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. Psychonomic Bulletin & Review, 23(4), 1015–1027. https://doi.org/10 .3758/s13423-015-0948-7

Received February 3, 2020 Revision received October 23, 2020

Accepted October 31, 2020