Journal of Experimental Psychology: Learning, Memory, and Cognition

Frequency Effects in Human Category Learning

Dong-Yu Yang and Darrell A. Worthy

Online First Publication, March 20, 2025. https://dx.doi.org/10.1037/xlm0001474

CITATION

Yang, D.-Y., & Worthy, D. A. (2025). Frequency effects in human category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://dx.doi.org/10.1037/xlm0001474

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001474

Frequency Effects in Human Category Learning

Dong-Yu Yang and Darrell A. Worthy
Department of Psychological and Brain Sciences, Texas A&M University

This study investigated the assumptions of prototype and exemplar models of human category learning, with a particular focus on the impact of category frequency. We used baseline and recency-weighted variants of prototype and exemplar models to examine the computational mechanisms underlying categorization decisions when one category was presented more frequently than the other. We employed extensive sets of stimuli derived from bivariate normal distributions and manipulated category frequency during training across four experiments using different category structures. In the transfer phases, participants classified novel stimuli. Across all studies, the results revealed a significant frequency effect, with participants showing a preference for categorizing novel items as members of the more frequently encountered category. This preference extended to transfer stimuli outside the trained region of the stimulus space. Model-based analyses indicated that the recency-weighted generalized context model exemplar model, which computes summed similarity via a Decay reinforcement learning rule, consistently outperformed other models in fitting the data and accurately reproducing the observed classification patterns across all experiments. Both prototype models failed to account for the observed frequency effects. While the baseline generalized context model was able to account for frequency effects, it did not capture recency effects. These findings suggest that relative category frequency influences human behavior when categorizing novel items. The computational modeling results revealed that evidence for categorization decisions is recency-weighted and accumulative rather than averaged.

Keywords: category learning, generalized context model, frequency effect, recency effect, reinforcement learning

Supplemental materials: https://doi.org/10.1037/xlm0001474.supp

Category learning is a crucial cognitive process that enables individuals to construct knowledge and organize their understanding of the world. The ability to establish and adopt categories is fundamental to the human population; it allowed our ancestors to identify and predict potential dangers in the environment and adapt to changing situations. As such, the study of category learning has been a central topic in cognitive psychology, neuroscience, and related fields. This has led to the development of numerous theoretical models to explain how individuals acquire and use category information. Understanding the mechanisms and processes underlying category learning can provide valuable insights into the nature of human cognition and its relation to behavior (Barsalou, 1985; Mareschal et al., 2010; Medin & Smith, 1984; Murphy & Medin, 1985).

Two major theories of human category learning—exemplar and prototype models—have been prominent for decades. Prototype models suggest people categorize new objects by comparing how similar the object is to the prototypes of each possible category (Mervis & Rosch, 1981; Rosch & Mervis, 1975). Posner and Keele (1968, 1970) found participants could extract the central tendency of attributes in the learning stage and were more likely to categorize the never-before-seen prototype as a category member in the testing stage (see also Reed, 1972; Smith & Minda, 2002). The prototype for one category is generally defined as the average or central tendency of each attribute across all category members. Unlike prototype models, exemplar models do not consider categories to be determined by comparing objects to a prototype. Exemplar models propose that all previously encountered exemplars are stored in memory. Categorization is performed by comparing the similarity between a novel object and all known exemplars within each possible category. If the object is most similar to the exemplars of one category, it will most likely be classified into that category (Brooks, 1978; Medin & Schaffer, 1978). The most representative exemplar-based theory is the generalized context model (GCM; Nosofsky, 1984).

In this article, we examine how well exemplar and prototype models can account for frequency effects in category learning, where participants show a bias toward a more frequently encountered category. Frequency effects have been observed in several aspects of cognitive processing. Estes's (1976a) seminal study revealed that, in a choice-prediction task, participants favored an option presented more frequently over an option presented less frequently. This preference

Andrew L. Cohen served as action editor.
Darrell A. Worthy https://orcid.org/0000-0001-5134-9745

Dong-Yu Yang is now at Department of Psychology, National Taiwan University.

The data and analysis code are available on the Open Science Framework at https://osf.io/kehrm/. The studies presented in this article were part of Dong-Yu Yang's doctoral dissertation at Texas A&M University.

Dong-Yu Yang played a lead role in data curation, formal analysis, and writing–original draft and an equal role in conceptualization, investigation, and methodology. Darrell A. Worthy played a lead role in supervision and writing–review and editing, a supporting role in formal analysis, and an equal role in conceptualization, investigation, and methodology.

Correspondence concerning this article should be addressed to Darrell A. Worthy, Department of Psychological and Brain Sciences, Texas A&M University, 4235 Texas A&M University, College Station, TX 77843-4235, United States. Email: worthyda@tamu.edu

can be observed even if the less-frequent option is associated with a higher reward probability (Don et al., 2019). In category learning, Homa et al. (1981, 1991) and Homa (1984) used the dot-pattern paradigm (Posner & Keele, 1968, 1970) and manipulated category size (or frequency), such as having categories contain three, six, or nine exemplars, in the learning phase in their experiments. They noted an increased prototype-enhancement effect with larger category sizes. In addition, Homa et al. (1981) revealed that classification accuracy for novel exemplars in the transfer phase was higher when these exemplars closely resembled trained ones. This old–new similarity effect diminished with increased category size. This was interpreted as evidence supporting the prototype-based models which suggests a prototype's representation becomes more dominant in categorization as category size expands. However, exemplar models can also account for the classic category size effects observed in dot-pattern paradigm studies (Nosofsky, 1988a; Shin & Nosofsky, 1992). Shin and Nosofsky (1992) not only replicated the prototype-enhancement and old–new similarity effects using the same paradigm but also revealed that the GCM offered a better explanation for those effects than the prototype model. The prototype-enhancement effect might arise because the prototype shares more similarity with all training exemplars (old distortions) than the novel exemplars (new distortions). This effect is augmented with larger category sizes due to more accumulated similarity representations. Conversely, the old–new similarity effect may decrease as category size grows because the similarity between new and old exemplars contributes less to the overall summed similarity (Hu & Nosofsky, 2022).

The present study focuses on how the relative frequency of each category affects the cognitive processes involved in category learning. We examine whether "frequency effects" exist in category learning, which we define as a bias toward classifying new stimuli as members of the more frequently encountered category. We also examine which models can account for such frequency effects. To provide theoretical grounding, we focus on the predictions made by variants of the GCM (Nosofsky, 1984, 1986, 1988b), as well as prototype models (Donald et al., 1973; Reed, 1972; Smith & Minda, 1998), in category learning tasks where the relative frequency of each category is manipulated. The GCM assumes the experienced exemplars are stored in memory with varying memory strengths. Memory strength might be influenced by the frequency or recency of the exemplar's presentation. Nosofsky (1988b) manipulated the exemplar's exposure frequency in the training phase of a category learning task and found evidence of a frequency effect. For instance, one exemplar was presented five times as often as any other exemplar within the target category in the high-frequency condition, whereas in the control condition, all exemplars were presented with equal frequency. Participants in the high-frequency condition had better categorization accuracy and rated the high-frequency exemplars as having higher category typicality than participants in the control condition. Furthermore, the increase was also detected for the stimuli which were near the high-frequency exemplars. Nosofsky (1988b, 1992) showed that the GCM can account for how exemplar frequency influences participants' categorization decisions. However, the memory strength parameters in the model were predetermined by experimental design and not learned by the model. This necessitates assigning a memory strength parameter to each category or exemplar, leading to a large number of parameters and questioning the GCM's parsimony.

In recent variants of the GCM, the memory for previously seen exemplars is thought to decay (McKinley & Nosofsky, 1995; Nosofsky et al., 1992, 2011; Zaki & Nosofsky, 2007), with more weight being given to more recent exemplars when classifying new stimuli. The idea of recency weighting suggests that more recent experiences are more influential in present cognitive decisions. Such recency effects have been well-established in memory studies and have also been identified in category learning. Past outcomes have been shown to influence participants' categorization behavior (Jones et al., 2006, 2013; Jones & Sieck, 2003; Little et al., 2016; Stewart et al., 2002), with participants being more likely to categorize novel items into the same category if they were similar to recent exemplars. Participants have been shown to give more weight to recent stimuli and incorporate recent experiences into subsequent categorization responses (Carvalho & Goldstone, 2015). These findings align with traditional incremental learning theories (Estes, 1950, 1957), such as the Delta reinforcement learning rule (Rescorla & Wagner, 1972; Widrow & Hoff, 1960).

However, while incremental learning, via the Delta rule, involves recency-weighted averaging, the Decay reinforcement learning rule (Erev & Roth, 1998) assumes recency-weighted accumulation of knowledge about the various decision-making alternatives. Because the Delta and Decay rules assume averaging and accumulation, respectively, in the present work, we used these rules to create recency-weighted versions of prototype and exemplar models. The recency-weighted prototype (RW-Prototype) model uses a Delta rule to compute the recency-weighted prototype for each category that is updated each time a new stimulus from a category is observed. This RW-Prototype model learns the recency-weighted average values for each stimulus dimension, and these values are used to represent the prototype of each category. In contrast, the RW-GCM model computes the recency-weighted summed similarity of the current stimulus to each previously encountered stimulus from each category. The summed similarity of each exemplar to the current stimulus decays more for exemplars encountered further in past, giving more recent exemplars greater weight. In addition to fitting recency-weighted variants of prototype and exemplar models, we also fit baseline versions.

These models are formalized and further described in the next section. As will be seen, the cumulative or summed representation of similarity in the exemplar models allows them to account for frequency effects much better than the prototype models. This is because summed similarity accumulates more for the more frequently encountered category. Based on the structure of the prototype models, we did not expect them to be able to account for frequency effects because the novel stimulus is only compared to the prototypes of each category, rather than to each exemplar. This process does not allow greater similarity to accumulate for the more frequently encountered category. Therefore, the prototype models are included mainly as examples of prominent category learning models that do not predict frequency effects.

Theoretical Models

Baseline GCM

In the GCM (Nosofsky, 1984, 1986, 1988b, 1992), each exemplar j is represented as a point in a multidimensional space. To compute the similarity between each exemplar j in the space and the current stimulus, i, the GCM uses the Minkowski power metric to calculate the distance d_{ij}:

$$d_{ij} = \left[\sum_{m=1}^{M} W_m |x_{im} - x_{jm}|^r\right]^{\frac{1}{r}}, \quad (1)$$

where the m is the dimension number from 1 to 2 because the stimuli varied along two dimensions in our category learning task. x_{im} denotes the value of stimulus i on dimension m, and x_{jm} denotes the value of stimulus j on dimension m. W_m is the attention weight granted to dimension m. Attention weights for the two dimensions are between 0 and 1 and are constrained to sum to 1. The stimuli in our study were lines varying in two perceptually separable dimensions of length and orientation. Hence, r is set to 1 in Equation 1 (Shepard, 1964). The distance d_{ij} is entered into an exponential-decay function (Shepard, 1987) to calculate the similarity s_{ij} between stimulus i and exemplar j:

$$s_{ij} = e^{-cd_{ij}}, \quad (2)$$

where c is the sensitivity parameter which represents the rate at which the similarity reduces with distance.

The baseline GCM assumes that people store information about all category exemplars they have encountered in memory. When classifying a new stimulus, the model compares its similarity to all stimuli of one category to its similarity to all stimuli from other categories. Stimuli are most likely to be classified as members of the most similar category. Thus, the probability that the stimulus i is classified into Category A is derived from:

$$P(A|i) = \frac{\sum_{j \in A} s_{ij}}{\sum_{j \in A} s_{ij} + \sum_{j \in B} s_{ij}}, \quad (3)$$

where Equation 3[1] denotes the summed similarity of stimulus i to stimuli from Category A, compared to the summed similarity of stimulus i to exemplars from both categories. The response bias and memory strength parameters of the GCM were estimated in previous studies (e.g., Nosofsky, 1988b). In this study, we assumed that the default preferences for each category were identical in our category learning task; thus, we did not include the response bias parameter in our model. Additionally, the inclusion of the response bias parameter would allow any of the models tested to account for frequency effects in a descriptive and arbitrary manner. For example, a higher response bias parameter for the more frequently encountered category could account for frequency effects. Our goal in the present study was to examine which models could account for frequency effects from their basic assumptions, rather than from the simple addition of a bias parameter.[2]

We also assumed that all stimuli (lines) have equal strength, and the memory strength parameter is not employed in the current model. In conclusion, based on the straightforward nature of the GCM mentioned, this article employs this simplified, baseline version of the GCM in further model-based analyses. Using simpler models, with fewer free parameters, will lead to more meaningful model fit comparisons, as each model is limited in the possible patterns of data that it could predict (Roberts & Pashler, 2000).

Baseline Prototype Model

Prototype models assume that people categorize new objects by comparing how similar the object is to the prototypes within related categories. The prototype of a category is usually described as a central tendency of all category members (Donald et al., 1973; Reed, 1972; Smith & Minda, 1998). The baseline Prototype model assumes that the prototype for Category A has the average values along each dimension, of all the k stimuli encountered from Category A. Formally, the average value for each m dimension for the prototype of Category A is updated each time a new Category A exemplar is seen according to:

$$P_{m,A}(k) = P_{m,A}(k-1) + \frac{1}{k} \cdot (x_m(k) - P_{m,A}(k-1)), \quad (4)$$

where $P_{m,A}(k)$ represents Category A's prototype value for dimension m after k exemplars have been shown from Category A, and we set the initial prototype dimension values ($P_{m,A}(0)$) to the average of each stimulus dimension across all stimuli, which approximates to the mean of the entire stimulus space. The prototype for Category B is updated in the same manner. Note that Equation 4 simply gives the average dimension values for each stimulus from a given category.

Minda and Smith (2001, 2002, 2011) suggested that the Prototype model best describes how people perform categorization. The model is closely related to the GCM. However, the stimuli are compared only to the category prototypes rather than all experienced exemplars. In Equation 5, d_{iP} represents the distance between stimulus i and prototype P:

$$d_{iP} = \left[\sum_{m=1}^{M} W_m |x_{im} - P_m|^r\right]^{\frac{1}{r}}. \quad (5)$$

As described in Equation 1, r is set to 1 due to the separable dimensions of stimuli employed in our category learning task, and W_m is the attention weight granted to dimension m.

Calculating the similarity s_{iP} between stimulus i and prototype P is identical to how the GCM computes similarity to each exemplar. The distance is converted into the similarity metric according to:

$$s_{iP} = e^{-cd_{iP}}, \quad (6)$$

where c is the sensitivity parameter. The stimulus is compared to the prototypes of each possible category using Equations 5 and 6.

[1] In later versions of the GCM, the response-scaling parameter (γ) has often been included as the exponent of the summed similarity for all compared categories (Ashby & Maddox, 1993; Nosofsky, 2011; Nosofsky & Zaki, 2002). We aimed to identify the simplest model to account for the frequency effect in categorization. Therefore, we did not include the response-scaling parameter in the present study to maintain simplicity. See also Footnote 1 in Smith (2014) which discusses how the addition of the gamma parameter allows for increased flexibility of the model which reduces its theoretical clarity.

[2] We note that many paradigms, such as signal detection, use response-bias parameters (Stephens et al., 2019); however, our focus was to examine how the models we tested could account for frequency effects without a response bias parameter.

The probability of stimulus i being categorized into Category A is then calculated by comparing the similarity to each prototype according to:

$$P(A|i) = \frac{s_{iP_A}}{s_{iP_A} + s_{iP_B}}. \tag{7}$$

Thus, the baseline Prototype model assumes that decisions are based on a similarity comparison to the average dimension values for each category.

Recency-Weighted GCM

The present study extends the GCM by introducing a recency-weighted GCM (RW-GCM) which uses a Decay reinforcement learning rule (Ahn et al., 2008; Don et al., 2019; Don & Worthy, 2022; Erev & Roth, 1998) to compute a decaying representation of the summed similarity of a given stimulus to the exemplars from each category. As proposed in McKinley and Nosofsky (1995, Equation 6), the memory strength of each exemplar likely decreases exponentially over time. The RW-GCM incorporates the Decay learning rule to compute accumulated similarity estimates. The overall similarity S_A for a novel stimulus to Category A at trial t is computed by applying the Decay rule to all the stimuli seen up to trial t. Beginning with the first exemplar j ($j = 1$), and updating through the most recent exemplar ($j = t$), the overall similarity of the current stimulus, i, to Category A, $S_{i,A}$, is updated according to:

$$S_{i,A}(j) = S_{i,A}(j-1) \cdot (1-\alpha) + s_{i,j} \cdot \delta_j, \text{for } j = 1,2,\ldots,t, \tag{8}$$

where $(1 - \alpha)$ represents the decay (forgetting) rate, and δ_j is a dummy variable that is 1 when exemplar j belongs to Category A, and 0 otherwise.[3] This equation is also used to compute the summed similarity to Category B. The summed similarity to each category decays each time a previous exemplar is compared to the current stimulus. Since Equation 8 begins with stimuli presented at the start of the experiment, the similarity of the current stimulus to the stimuli seen earlier in the task will decay the most. Because similarity is only added, or incremented to the summed similarity for the correct category$(s_{i,j} \cdot \delta_j)$, summed similarity will accumulate more for the category that is presented more frequently. The similarity to each j stimulus is computed using Equations 1 and 2 from the baseline GCM. A similar concept was proposed by McKinley and Nosofsky (1995), which developed a recency-sensitive variant of GCM, and the assumption of memory decay for exemplars has been prominent in other works as well (Nosofsky et al., 2011; Zaki & Nosofsky, 2007). Note that when $\alpha = 0$ in Equation 8, the RW-GCM is identical to the baseline GCM, and when $\alpha > 0$ the model assumes recency weighting. Thus, the baseline GCM is nested within the RW-GCM as a special case.

Recency-Weighted Prototype Model

Equation 4 is a special case of a Delta learning rule, which tracks the incremental average of a series of values (Rescorla & Wagner, 1972; Widrow & Hoff, 1960). That equation gives equal weight to all previously seen exemplars, but for the RW-Prototype model, we used the more general form of the Delta learning rule that gives greater weight to more recent information.[4] Each time a new k exemplar from Category A is seen, the Category A prototype for each m dimension is updated according to:

$$P_{m,A}(k) = P_{m,A}(k-1) + \alpha \cdot (x_m(k) - P_{m,A}(k-1)). \tag{9}$$

Here, α is a free parameter that is interpreted in the same manner as α in Equation 8 for the RW-GCM model, with higher values indicating greater weight to more recent information. In this model, we initialized the prototype dimension values ($P_{m,A}(0)$) at the average values for each stimulus dimension, across all stimuli, or at the approximate average of the stimulus space. This model essentially assumes that the prototype for each category is pulled toward each new stimulus that is encountered, with the effect or "pull" of each stimulus diminishing as new stimuli are observed. It is also important to note that when $\alpha = 1/k$, the RW-Prototype model is equivalent to the baseline Prototype model. When $\alpha > 1/k$, then the model assumes recency weighting of previous stimuli, and when $\alpha < 1/k$, the model assumes more primacy weighting than the baseline GCM, or a bias toward the initial starting values.

The probability of stimulus i being categorized into Category A is calculated on each trial according to Equation 7. For all models, the probability of selecting Category B can be given by subtracting the probability of an A response from 1.

To summarize, the four models we compare in the present work have free parameters for sensitivity (c) and the attention weight (W). This means the baseline GCM and Prototype models have two free parameters. The RW-GCM and RW-Prototype models also include additional recency parameters (α) to represent the learning and decay rates, which capture recency effects.

Aims of the Present Study

The primary objective of this study is to explore the influence of category frequency on the category learning process. In many previous studies (e.g., Nosofsky, 1988b), the number of exemplars has been relatively small, and researchers often predetermined categories arbitrarily. These studies generally reported high categorization accuracy during training, suggesting that participants learned the exemplars effectively. Although this setting may be consistent with the GCM's assumptions about exemplar memory, its generalizability to everyday life outside the laboratory may be limited. To address this, we applied a multidimensional sampling

[3] Because time progresses and exemplar memory decays continuously, the recursive form of the RW-GCM computes summed recency-decayed similarity of a novel stimulus to category exemplars at each trial. This contrasts with the baseline GCM, which simply sums similarity across all exemplars without considering recency effects. Specifically, at trial t, the summed similarity for the first j stimuli ($S_{i,A}(j)$) decays further by trial $t+1$, as older exemplars are given less weight due to the decay parameter. Thus, the probability of categorizing a stimulus i into Category A is calculated as: $P(A|i) = S_{i,A}/(S_{i,A} + S_{i,B})$, which fits within the context of the RW-GCM.

[4] A GCM variant incorporating the Delta rule (Delta-GCM), which we present in the Supplemental Materials, computes recency-weighted averaged similarity to exemplars from each category, making it a plausible model for investigating recency effects in categorization. However, due to its averaging process, the Delta-GCM is not expected to predict frequency effects effectively. This limitation arises because the model does not accumulate similarity across exemplars but instead averages them, which prevents the summed similarity from reflecting category frequency. Simulations for the Delta-GCM under the conditions of Experiment 1a are presented in Supplemental Figure S18, p. 26.

approach, which has been implemented by Ashby and Gott (1988), to generate a large set of unique stimuli for our category learning task. Given the extensive number of stimuli, it is impractical for participants to memorize each one, providing a robust test of the GCM's generalizability.

In addition, we aimed to examine whether our proposed variants incorporating reinforcement learning mechanisms, specifically recency-weighted variants of both exemplar and prototype models, could more effectively capture frequency effects in categorization. We hypothesized that the GCM variants would better capture frequency effects than the Prototype models due to GCM's property of summing similarity over categorizations. We also predicted that the RW-GCM which incorporates the Decay rule would better account for frequency effects than the baseline GCM model because it takes recency effects into account. The Prototype models will likely be unable to account for frequency effects because they compare new stimuli to a single prototype rather than accumulating similarity.

To summarize, this study compares four models—baseline GCM, RW-GCM, baseline Prototype, and RW-Prototype—on their ability to account for frequency effects in categorization. The models are tested using hundreds of unique stimuli sampled from trained and untrained regions of the stimulus space, ensuring decisions are based on general category knowledge rather than memorization of specific exemplars. We also use three distinct category structures as well as novel transfer stimuli to test the generalizability of our findings. By incorporating insights from reinforcement learning and systematically varying category frequency, this work aimed to shed light on the mechanisms driving frequency effects in category learning and the models best suited to explain them.

General Method

This study comprises four experiments that each have the same basic procedures, except the stimuli used in each differ. To enhance clarity, we describe the basic methods used across all experiments in this section. Unique methods specific to each experiment are presented in their respective method sections.

Participants

All participants were undergraduate students at Texas A&M University who participated in the studies for partial credit as a requirement for a course offered by the Department of Psychological and Brain Sciences. Each participant was only allowed to participate in one of the conditions among four studies. Participants gave informed consent before the experiment. The experimental methods and procedures were approved by the Texas A&M University Internal Review Board.

Apparatus and Procedure

The category learning task was performed in a Chrome web browser. The Hypertext Markup Language-based webpage which presented the experimental task was programmed using JavaScript and the jsPsych JavaScript library (de Leeuw, 2015). Participants completed the study using a personal computer with a 1,920 × 1,080 screen resolution in the laboratory.

Participants read the consent form and completed a brief demographic survey; they were then assigned to one of the conditions under one of the studies. They were instructed to operate the assigned category learning task using a computer mouse. Upon completing the task, they read the debriefing information on the screen and received course credit for their participation.

Categorical Stimuli

Stimuli consisted of lines that varied in both dimensions of length and orientation, similar to those which have been applied in prior category learning studies (Cornwall et al., 2022; Filoteo et al., 2010). Sample line stimuli used in the present study are presented in the Supplemental Materials. We used a randomization technique introduced by Ashby and Gott (1988) to sample two stimuli sets for two categories in each of the four studies. Each random sample (x, y) was converted to a stimulus by deriving the length ($=x$, pixel) and the orientation ($=y \cdot \pi/500$, radians).

Trial Procedure

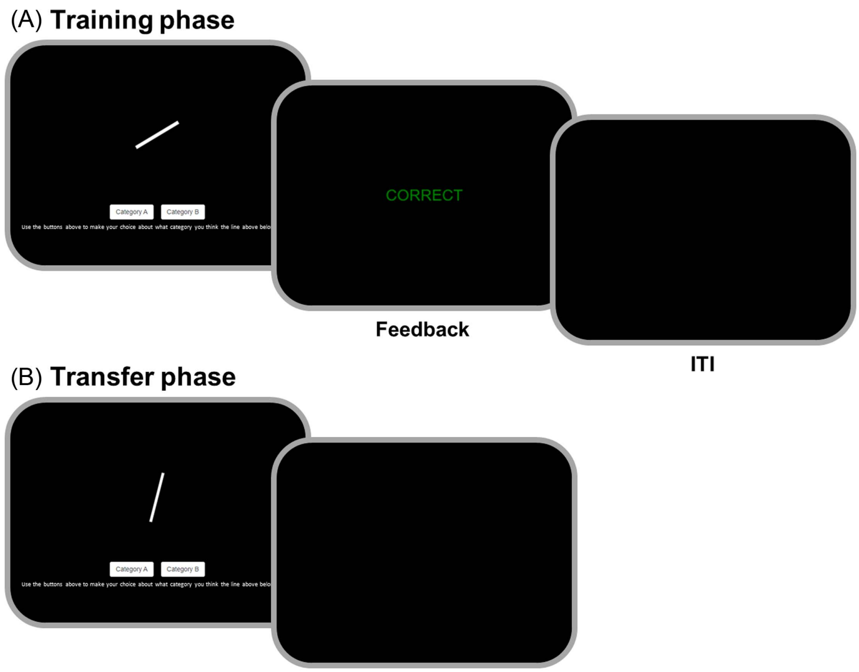
Figure 1 illustrates the trial timeline for our category learning task. Each condition consisted of two 200-trial blocks. Participants were instructed to categorize the lines into either Category A or B. On each trial, a line was presented and remained on the screen until the participant made a choice by dragging the mouse and clicking one of the buttons labeled "Category A" or "Category B" appearing below the line on the screen. In the training phase, corrective feedback was provided for 1 s after the response. The next trial was initiated following a 0.5-s intertrial interval. In the transfer phase, no corrective feedback was provided, and the next trial began after a 1.05-s intertrial interval. Choices were self-paced.

Category Structures

Three different category structures were applied to define category memberships in our four experiments. During the training phase, participants were exposed to stimuli that fell into each category. Experiments 1a and 1b applied an information-integration (II) structure used in prior studies (Ashby & Maddox, 2005; Daniel & Pollmann, 2010). The II structure requires participants to integrate two dimensions of information to categorize the stimuli. This task requires a more implicit or procedural form of learning because the optimal strategy is difficult to verbalize. The left panel of Figure 2A demonstrates the structure used in Experiments 1a and 1b, which was first seen in Ashby and Gott (1988). In Experiment 2, we applied a conjunctive (CJ) rule category structure, where the optimal boundaries are orthogonal to the axes of the dimensions. This enables the optimal strategy to be more easily verbalized (Ashby et al., 1998; Ashby & Maddox, 2005). For example, in the middle panel of Figure 2A, one could easily find the rule to discriminate between two categories by such a description: Short and steep lines are classified in one category, and all others are in the other category. In Experiment 3, we applied a unidimensional (UD) structure in which the category membership was determined by one dimension (Zeithamova & Maddox, 2006). The categories in the UD structure are easy to identify compared to the II and CJ structures because people can attend to only one dimension to make the decision for categorization. The right panel of Figure 2A illustrates the UD structure in Experiment 3.

Figure 1

Trial Timeline of (A) Training Phase and (B) Transfer Phase of the Category Learning Task



Note. Participants made categorization decisions by clicking the corresponding buttons for Category A or B at each trial. Following the participant's choice, text feedback was presented for 1 s (training only). There was then a blank screen shown in the intertrial interval (training: 0.5 s; transfer: 1.05 s). ITI = intertrial interval. See the online article for the color version of this figure.

In the transfer phase, novel stimuli were presented to participants. In Experiment 1a, the identical structure as in the training phase was combined with stimuli from two untrained regions of the stimulus space to create the category structure for the transfer phase (see Figure 2B). Starting from Experiment 1b, the transfer phase employed a grid structure consisting of 121 novel stimuli evenly distributed across the entire stimulus space based on two dimensions (see Figure 2C).

Category Frequency

In the training phase, participants were presented with one of 400 unique lines that varied in length and orientation on each trial in a randomized order. The frequency of the two categories was manipulated in the training phase. In the A-Frequent condition, the number of trials for Category A is roughly two times greater than Category B (266:134). Conversely, Category B has two times the number of trials as Category A in the B-Frequent condition. In the Even-Frequent condition, the stimuli for the two categories are equal (200:200).

Data Analyses

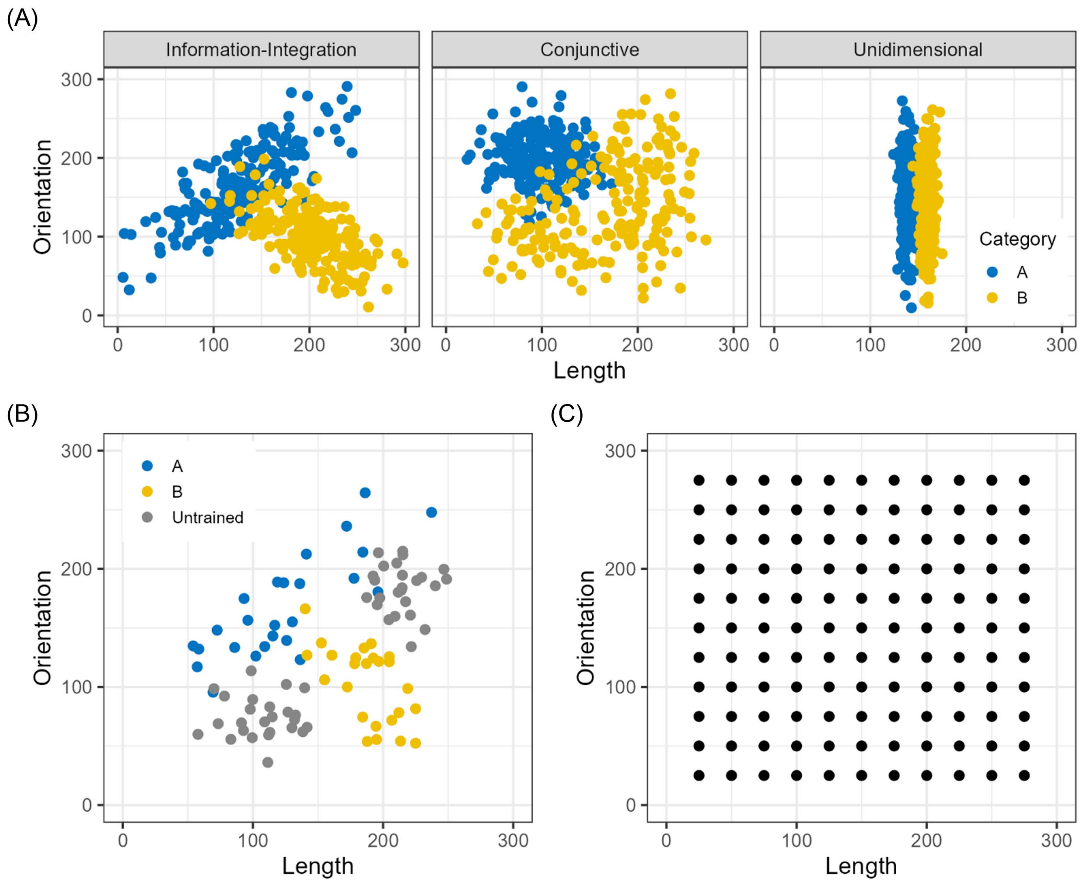
We analyzed the data in two separate phases. During the training phase, we specifically looked at the proportion of Category A choices within each category. We divided the 400 trials into four 100-trial blocks for analysis. Using JASP Version 0.18.3 (JASP Team, 2023), we conducted Bayesian repeated measures analyses of variance (ANOVAs) to investigate the impact of category frequency on the proportion of Category A choices across blocks for the exemplars from each category. We performed the same analysis for the transfer phase. Additionally, we analyzed the proportion of correct responses in the training phase and the response times for categorization decisions in both phases. Below, we focus primarily on the proportion of Category A choices in each condition, as this metric is most informative regarding frequency effects. However, more detailed results for each experiment can be found in the Supplemental Materials.

Model-Based Analyses

In each experiment, we employed a multifaceted approach to model-based analyses. We first conducted a priori simulations using the baseline GCM, baseline Prototype, RW-GCM, and RW-Prototype models to predict behavioral patterns in categorization tasks. We applied the Maximum Likelihood Method for model fitting and assessed model performance using the Bayesian information criterion (BIC). The best-fitting parameters for each model were then used in post hoc simulations to reproduce each participant's categorization sequence. All simulations and model fits assumed that learning only occurred during training and that there was no decay of information across the transfer period. In the transfer phase, the dimension values for each prototype were fixed for the prototype models. For the exemplar models, summed similarity to each transfer stimulus was computed using only stimuli encountered during training. The data and analysis code are available at https://osf.io/kehrm/

Figure 2

Overview of Category Structures, Stimulus Spaces, and Novel Stimuli Distributions Across Experiments 1, 2, and 3



Note. (A) Three category structures—information-integration, conjunctive, and unidimensional—were employed in Experiments 1, 2, and 3, respectively. (B) The stimulus space for the transfer phase of Experiment 1a. Category A (25 trials) and Category B (25 trials) stimuli were sampled from the same distributions as in the training phase. Untrained stimuli (gray dots) were not employed in the training phase. Two untrained regions, located in the top right and bottom left of the space, comprised 25 trials each. (C) A total of 121 novel stimuli were evenly distributed across the entire space, forming a grid structure used in Experiments 1b, 2, and 3 for the transfer phase. See the online article for the color version of this figure.

A Priori Simulations

We conducted a priori simulations to forecast the impact of frequency on categorization accuracy and the proportion of Category A choices in both the training and transfer phases across different frequency conditions. A total of 1,000 simulated agents performed the task in each training condition for each model, with their parameters randomly sampled from a uniform distribution using the following parameter bounds: GCM: $W \sim U(0, 1)$; $c \sim U(0, 0.5)$; Prototype: $W \sim U(0, 1)$; $c \sim U(0, 0.5)$; RW-GCM: $W \sim U(0, 1)$; $c \sim U(0, 0.5)$; $(1 - \alpha) \sim U(0, 1)$; RW-Prototype: $W \sim U(0, 1)$; $c \sim U(0, 0.5)$; $\alpha \sim U(0, 1)$. The simulation results were averaged across the parameter combinations.

Model Fitting

We used *optim* in R (R Core Team, 2024) to estimate the best-fitting parameters for each model. Each participant's data were fit individually by maximizing the log-likelihood of each model's prediction for participants' responses on each trial. To compare the fit of different models, we computed the BIC (Schwarz, 1978) which is adjusted based on each model's complexity (i.e., the number of parameters):

$$\mathrm{BIC} = -2LL + g \cdot \ln(N - 1), \tag{10}$$

where $N - 1$ represents the total number of trials minus 1,[5] and g represents the number of free parameters. Lower BIC scores indicate a better model fit. To assess the relative improvement of one model over another, we evaluated the difference in BIC values (ΔBIC) between the models. The BIC differences were then converted into Bayes factors, $\mathrm{BF} = \exp(\Delta\mathrm{BIC}/2)$, to indicate

[5] The likelihood for the first trial is not included because the choice is assumed to be random.

the extent to which the best-fitting model was superior to each other model. A BF value between 1 and 3 is considered weak evidence in favor of the best-fitting model, a range of 3–20 is viewed as positive evidence, and a range of 20–150 represents strong evidence in favor of the best-fitting model (Masson, 2011; Wagenmakers, 2007).

Post Hoc Simulations

The best-fitting parameters for each participant were used as input for the models of interest to run 100 simulations for each participant. The resulting simulation data were then averaged across all simulations for each participant to generate the average predicted probability of selecting each option for each stimulus. The empirical data were processed similarly to compare with each model's predicted proportion of Category A choices for each stimulus. The mean square deviation (MSD) score was calculated using the following formula:

$$\mathrm{MSD} = \frac{1}{n}\sum_{j=1}^{n}\left(\bar{P}_{\exp,A}(j) - \bar{P}_{\mathrm{sim},A}(j)\right)^2, \tag{11}$$

where n is the number of stimuli, $\bar{P}_{\exp,A}(j)$ is the average proportion of Category A choices for stimulus j across participants, and $\bar{P}_{\mathrm{sim},A}(j)$ is the average proportion of Category A choices for stimulus j across simulated agents. The MSD reveals the deviation of the simulated data from the observed data. The lower the score, the higher the model's accuracy in reproducing participants' actual responses.

Experiment 1

Experiment 1 aimed to investigate the frequency effect during category learning. We used the category structure illustrated in Figure 2A. Experiment 1 consisted of two subexperiments: 1a and 1b. The only distinction between Experiments 1a and 1b was the category structure used in the transfer phase. Below, we present the a priori simulation results before the Method section to outline the predictions for each model.

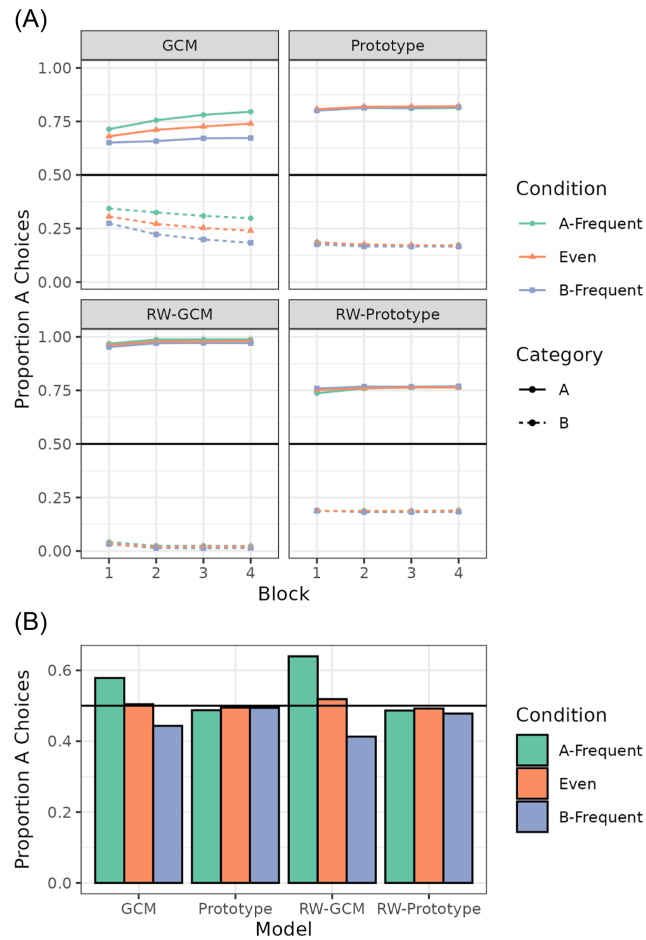
Experiment 1a

Figure 3A shows the average proportion of Category A choices for Category A and B stimuli during training. Solid lines show the proportion of A choices for Category A exemplars, and dashed lines show the proportion of A choices for Category B stimuli. The simulations showed that the baseline GCM and the RW-GCM predicted frequency effects. Note that the baseline GCM predicts frequency effects even for stimuli that are from the less-frequent category. More A choices are predicted in the A-Frequent condition for both Category A and B stimuli, and the same pattern is predicted for B choices in the B-Frequent condition. The Prototype models did not predict frequency effects.

We also found that the simulated accuracy rates were extremely high for the RW-GCM model. This is likely because summed similarity to stimuli from each category exponentially decays as newer stimuli are processed, leading to larger differences in summed similarity between the correct and incorrect categories, compared to the baseline GCM where there is no exponential decay of similarity.

Figure 3

Proportion of Category A Choices in a Priori Simulations for Experiment 1a



Note. (A) The mean proportion of Category A choices averaged over each 100-trial block for exemplars from each category (Category A: solid line; Category B: dashed line) across three conditions in the simulated training phase. (B) The mean proportion of Category A choices for each condition in the simulated transfer phase. The black horizontal line represents a proportion of 0.5. RW = recency-weighted; GCM = generalized context model. See the online article for the color version of this figure.

The Luce Choice rule in Equation 3, which determines the probability of selecting each category, is influenced by the *ratio* of summed similarity to each category (Worthy et al., 2008), rather than the absolute difference in similarity. High decay rates (α) in the RW-GCM lead to much larger ratios between the summed similarity to the correct versus incorrect categories, compared to lower decay rates. Supplemental Figures S15–S17 show similar plots using specific parameter values for the decay parameter, with the sensitivity parameter set to 0.05. The accuracy curves for the lowest decay rate of 0.1 are very similar to the baseline GCM, and they do not depict the very high accuracy shown in Figure 3A. However, the predicted accuracy rates are very high when the decay rate is set to 0.5 or 0.9. In models fit to our data reported below, we generally observe decay rates much lower than 0.5, and usually lower than 0.05 for the RW-GCM, which allows the model to predict reasonable accuracy curves that are similar to those from the baseline GCM.[6]

In the transfer phase, strong frequency effects were predicted for the baseline GCM and the RW-GCM but not for the variants of Prototype models. The RW-GCM also appeared to predict a larger frequency effect than the baseline GCM (Figure 3B). These simulation results provided clear hypotheses for the experiments.[7] Models that incorporate a cumulative similarity metric (GCM and RW-GCM) predict strong frequency effects, while models that compute an averaged similarity metric predict no effect of category frequency (Prototype and RW-Prototype).

Method

Participants

A total of 175 students at Texas A&M University (101 female participants, one preferred not to respond) with a mean age of 18.73 (SD = 1.03; five participants did not report their age) participated in experiments to partially fulfill a course requirement. Participants were randomly assigned to one of three experimental conditions. Participants who did not achieve at least 50% accuracy in the last 100 trials of the training phase were not included in further analyses. This was to ensure participants achieved the required level of learning during the training phase. The final sample sizes per condition were 58 in the A-Frequent condition (one dropped), 58 in the B-Frequent condition, and 57 in the Even-Frequent condition (one dropped).

Design

Experiment 1 employed two II categories in the current category learning task shown in the left panel of Figure 2A. For the transfer phase, an additional 100 unique lines were presented from the transfer structure shown in Figure 2B. The detailed category distribution parameters for both categories are provided in Table 1.

Procedure

The procedures for the training phase and the transfer phase are described in the General Method section.

Table 1
Category Distribution Parameters for the Length and Orientation Dimensions in the II Category Structure Used in Experiment 1

	Parameter				
Stimulus space	μ_l	μ_o	σ_l	σ_o	r
Category					
A	130	170	45	45	.75
B	200	100	35	35	−.60
Untrained spaces					
1	115	85	20	20	0
2	215	185	20	20	0

Note. Dimensions are in arbitrary units. Please refer to Figure 2A for the scatterplot. The subscripts "l" and "o" denote length and orientation, respectively. II = information-integration.

Results

Behavioral Analyses

The results from the training phase of Experiment 1a are presented in Figure 4A. The results revealed a higher proportion of Category A choices in the A-Frequent condition than in the Even- and the B-Frequent conditions for both Category A and B stimuli. In other words, participants preferred categorizing the stimuli into categories that were presented more frequently in the training phase. This frequency effect was observed for stimuli from both categories.

We performed 3 (Condition) × 4 (Block) Bayesian Repeated Measures ANOVAs[8] to assess the effect of frequency condition on the proportion of Category A choices over blocks in the categories from which the stimuli were presented. For Category A, as displayed in the upper part of Figure 4A, the frequency effect was little or none in the early blocks and enlarged in the later blocks. The analysis revealed that the data most support the model with the Block × Condition interaction ($BF_{10} = 2.39 \times 10^{33}$). For Category B, the data support the best model that includes the main effect of Block and Condition without the interaction ($BF_{10} = 1.07 \times 10^{30}$), suggesting that the frequency effect was consistent over blocks, as presented in the bottom part of Figure 4A.

In the transfer phase, we observed the same frequency effect on the proportion of Category A choices (Figure 4B). We used a one-way Bayesian ANOVA to examine the differences among three category frequency conditions. The data strongly supported the model that includes the main effect of the condition of category frequency ($BF_{10} = 1.25 \times 10^{25}$).[9] Post hoc comparisons indicated that more Category A choices were detected in the A-Frequent condition than in the Even condition ($BF_{10} = 1.16 \times 10^{6}$) and the B-Frequent condition ($BF_{10} = 1.23 \times 10^{23}$). There were also more A choices in the Even condition than in the B-Frequent condition ($BF_{10} = 1.94 \times 10^{8}$).

We next examined the frequency effect during the transfer phase for stimuli from different areas of the stimulus space. A 3 (Space) × 3 (Condition) Bayesian Repeated Measures ANOVA indicated strong support for the model encompassing the Space × Condition interaction ($BF_{10} = 1.31 \times 10^{187}$). Simple main effects were

[6] Our a priori simulations are considered *data-uninformed* (Wagenmakers et al., 2004) because we did not make any assumptions about the parameters that would most likely align with human behavior. In contrast, our post hoc simulations are considered *data-informed.* We felt that performing both types of simulations would help us best understand the pattern of each model's predictions. The simulations shown in the Supplemental Material use fixed parameter estimates that were thought to lead to reasonable predictions.

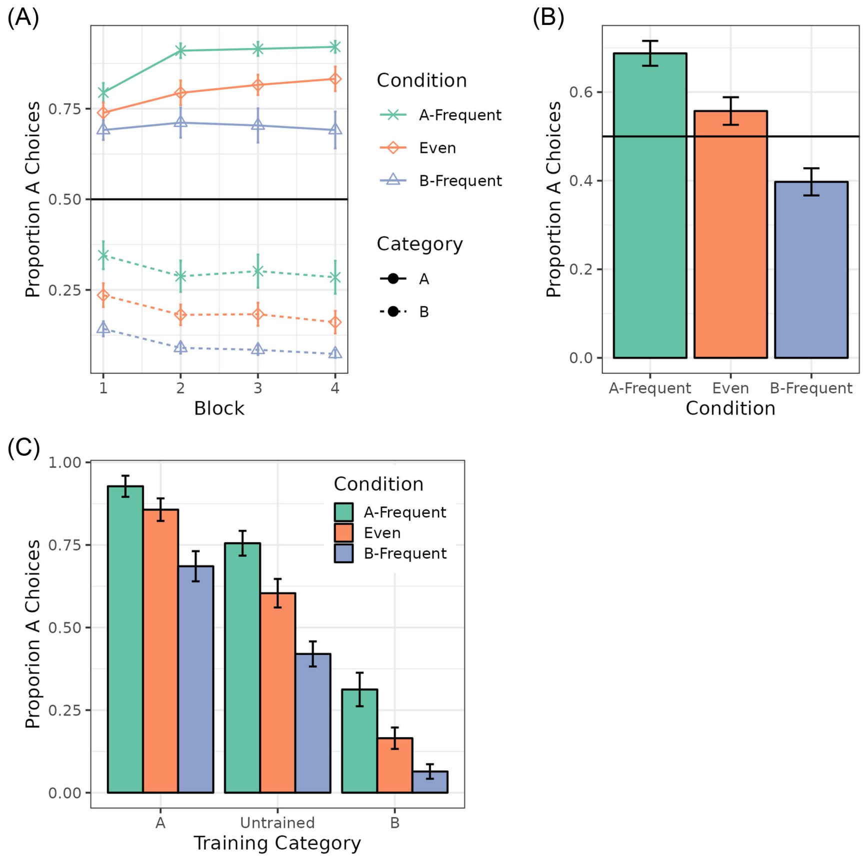
[7] We also conducted a priori simulations with fixed values for c (0.05) and W (0.5) and three different values of α (0.1, 0.5, 0.9) to examine how these combinations of parameters influence predictions for frequency effects. The results are presented in the Supplemental Materials (pp. 23–25). The baseline GCM and RW-GCM can predict frequency effects at all three α values used in these simulations.

[8] We planned to run comparisons between conditions for both category A and B stimuli. We attempted to run 2 (Category) × 3 (Condition) × 4 (Block) Bayesian Repeated Measures ANOVAs; however, the statistical program (JASP) repeatedly crashed when attempting to estimate this model. Therefore, we report the results of analyses for stimuli from each category separately.

[9] For consistent statistical reporting, we have presented the Bayes factors provided in JASP for the post hoc tests in this article. It is important to note that the post hoc tests were not corrected for multiple comparisons.

Figure 4

Mean Proportion of Category A Choices During Training and Transfer Phases Across Three Conditions in Experiment 1a



Note. (A) The mean proportion of Category A choices during the training phase is shown for exemplars from two categories (Category A: solid line; Category B: dashed line), averaged for each 100-trial block across three conditions. (B) The mean proportion of Category A choices during the transfer phase is shown for each condition. The black horizontal line represents a proportion of 0.5. (C) The mean proportion of Category A choices is shown across three transfer spaces. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

examined for each training space. There was a strong main effect for stimuli within Category A space ($BF_{10} = 1.23 \times 10^{13}$). Post hoc tests showed a higher proportion of Category A choices in the A-Frequent condition than in the Even condition ($BF_{10} = 11$). Strong evidence also supported a higher proportion of A choices in the Even condition relative to the B-Frequent condition ($BF_{10} = 440{,}349$).

In the Category B space, the main effect of frequency received robust support ($BF_{10} = 4.23 \times 10^{13}$). Post hoc analysis revealed that the A-Frequent condition surpassed the Even condition in the proportion of Category A choices ($BF_{10} = 4{,}503$), and the Even condition was higher than the B-Frequent condition ($BF_{10} = 13{,}325$).

In the untrained space, there was again a strong main effect of frequency ($BF_{10} = 2.15 \times 10^{20}$). Post hoc tests indicated that participants were more likely to choose Category A in the A-Frequent condition than in the Even condition ($BF_{10} = 23{,}584$) and more so in the Even condition than in the B-Frequent condition ($BF_{10} = 2.73 \times 10^{6}$).

Model-Based Analyses

The average best-fitting parameter estimates are presented in Supplemental Table S1. Table 2 reveals the mean BIC, ΔBIC, and Bayes factors across participants for each model for each condition. The RW-GCM provided the best fit to the data in all conditions. Bayes factors results showed strong evidence supporting the RW-GCM fitting better than the baseline GCM and the Prototype models. Consequently, we conducted a Bayesian one-way ANOVA to compare the best-fitting parameter estimates for the RW-GCM across each condition. No evidence supported an effect of condition for any of the parameters, sensitivity (c; $BF_{10} = 0.60$), attention weight to the dimension of length (W; $BF_{10} = 0.20$), decay rate ($1 - \alpha$; $BF_{10} = 0.82$).

We next conducted post hoc simulations to evaluate the models on their ability to reproduce the actual behavioral patterns seen in participants. Table 2 shows the MSD scores from each model in each of the three conditions. The RW-GCM outperformed others

Table 2
Mean BIC, ΔBIC, Bayes Factors, and MSD for Each Model in Experiment 1a

Condition	Model	BIC	ΔBIC	log (BF)	MSD
A-Frequent	GCM	389	22	4.78	0.0068
	Prototype	456	89	19.33	0.0323
	RW-GCM	**367**			**0.0053**
	RW-Prototype	462	95	20.63	0.0343
Even	GCM	427	22	4.78	0.0071
	Prototype	459	54	11.73	0.0214
	RW-GCM	**405**			**0.0048**
	RW-Prototype	457	52	11.29	0.0207
B-Frequent	GCM	376	20	4.34	0.0106
	Prototype	464	108	23.45	0.0428
	RW-GCM	**356**			**0.0081**
	RW-Prototype	466	110	23.89	0.0390

Note. ΔBIC represents the BIC difference between each model and the best model shown in bold. MSD was calculated for all trials in both the training and transfer phases. BF represents the Bayes factor, calculated as exp(ΔBIC/2). The smallest value within each condition is presented in bold. For brevity, we report the logarithm-transformed value, log (BF). BIC = Bayesian information criterion; MSD = mean square deviation; RW = recency-weighted; GCM = generalized context model; exp = exponential.

and produced the lowest MSD score in all conditions. The baseline GCM and the RW-GCM most accurately reproduced the data patterns in the training phase, as depicted in Figure 5A. For the transfer phase, Figure 5B illustrates that the baseline GCM and the RW-GCM reproduced the frequency effects observed in the empirical data. The baseline Prototype and RW-Prototype models failed to generate this pattern. Figure 6 shows the scatterplot detailing the associations between simulated and observed Category A choice proportions in the transfer phase between the three frequency conditions at the individual stimulus level. The tight clustering around the line of perfect correlation indicates that the GCM variants have greater predictive accuracy for the empirical data across the different conditions. On the other hand, the Prototype variants exhibit greater variability and less agreement with the empirical data, indicating a lower degree of predictive accuracy for the observed percentages of Category A choices.

Experiment 1b

In Experiment 1a, we successfully induced participants to make more categorization decisions to the high-frequency category by manipulating category frequency in the training phase. A frequency effect was observed in the transfer phase, wherein participants tended to categorize novel stimuli into the category to which they were more often exposed during training. Moreover, this categorization preference was also observed with novel stimuli generated from untrained areas of the stimulus space. In Experiment 1b, we sought to investigate how the frequency effect generalizes to a wider transfer space. Experiment 1b used the same category structure in the training phase as in Experiment 1a. However, a new grid structure, which included a wider area than that for the trained categories, was used in the transfer phase. This new transfer structure is the only difference in design from Experiment 1a. As described in the General Method section, Figure 2C illustrates the grid structure applied in the transfer phase.

As in Experiment 1a, four categorization models were employed to simulate the category learning task. Supplemental Figure S3B[10] indicates that two GCM variants predicted the frequency effects on the proportion of Category A choices in the grid structure transfer phase. Compared with Experiment 1a, most models predicted an overall bias toward Category A in the transfer phase. This is likely due to the nature of the category structure used, where Category A is more broadly distributed across the stimulus space than Category B, which is more tightly clustered. The baseline GCM and RW-GCM, which have the similarity-cumulating property, both predicted frequency effects in the transfer phase, with the RW-GCM predicting the greatest differences in Category A choices across conditions.

Method

Participants

A total of 159 students at Texas A&M University (117 female participants, one preferred not to respond) with a mean age of 18.57 ($SD = 2.28$; three participants did not report their age) participated in the experiment to partially fulfill a course requirement. Participants were randomly assigned to one of three experimental conditions. Participants who did not achieve at least 50% accuracy in the last 100 trials of the training phase were not included in further analyses. This was to ensure participants achieved the required level of learning for categories of stimuli. The sample sizes per condition were 51 in the A-Frequent condition (one dropped), 55 in the B-Frequent condition, and 52 in the Even-Frequent condition.

Design

Experiment 1b employed the identical II structure and the same manipulation of category frequency as Experiment 1a for the training phase. In the transfer phase, unlike the structure described in the transfer phase of Experiment 1a, a grid structure was employed (see Figure 2C).

Procedure

The task procedure in Experiment 1b was identical to that described in the General Method section.

Results

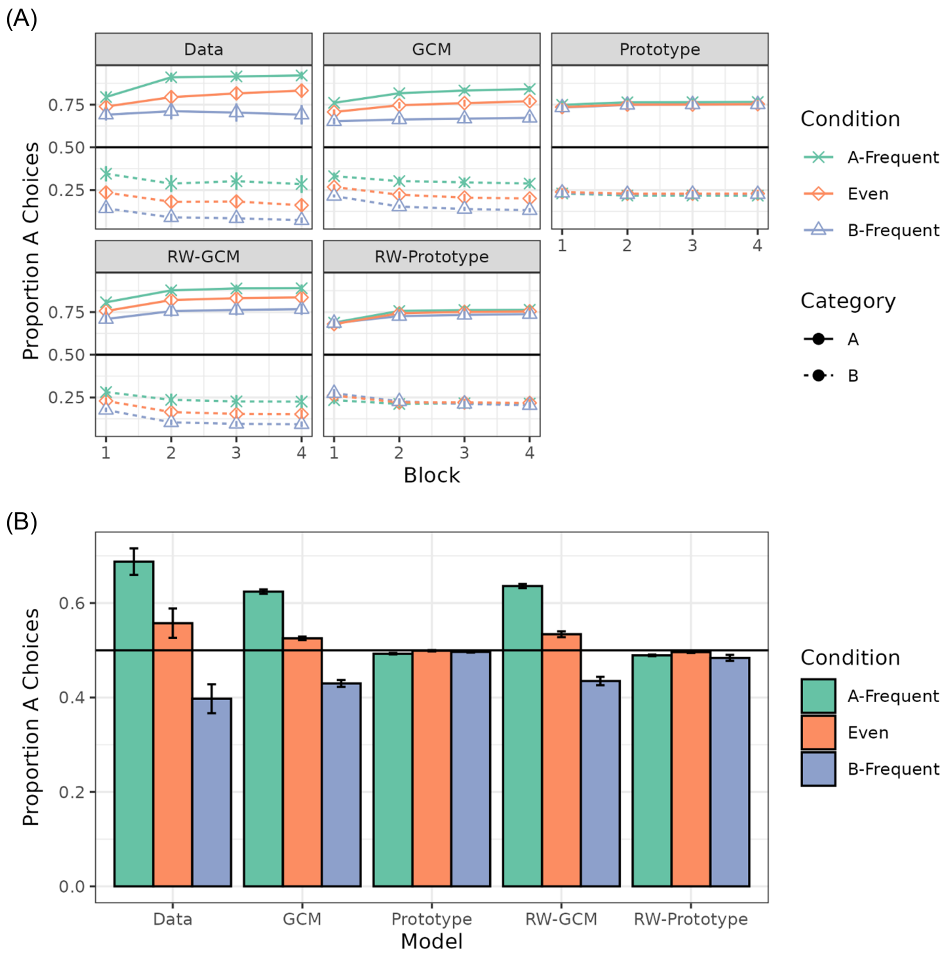
Behavioral Analyses

In the training phase, as in Experiment 1a, a frequency effect was observed. Participants selected the category that occurred more frequently for Category A and Category B stimuli. Figure 7A shows the proportion of Category A choices for Category A and B stimuli for each frequency condition. We again conducted 3 (Condition) × 4 (Block) Bayesian Repeated Measures ANOVAs to test the frequency effect over blocks in the categories from which the stimuli were presented. For Category A, the analysis revealed that the data most strongly support the model that includes the Block × Condition interaction ($BF_{10} = 1.27 \times 10^{16}$), suggesting that the frequency effect is unequal over blocks. We also observed

[10] For brevity, we report the a priori and post hoc simulation for Experiment 1b and subsequent experiments in the Supplemental Materials, as the simulation patterns are quite similar in the following experiments.

Figure 5

Proportion of Category A Choices in Experiment 1a: Empirical Data Versus Post Hoc Simulations



Note. The figure presents the mean proportion of Category A choices for simulated and observed data for each condition. (A) For the training phase, the top-left panel displays the observed data. The remaining panels show the results from model simulations. (B) For the transfer phase, the left column displays the observed data. The remaining columns show model simulation results. The black horizontal line represents a proportion of 0.5. Error bars represent 95% confidence intervals. RW = recency-weighted; GCM = generalized context model. See the online article for the color version of this figure.

that the effect is smaller between the Even and B-Frequent conditions than between the A-Frequent conditions and the Even conditions. The data for Category B indicate that the frequency effect is consistent across blocks; the best-supported model included the main effects of Block and Condition without the interaction ($BF_{10} = 8.02 \times 10^{28}$).

We observed similar frequency effects in the transfer phase (Figure 7B). Participants in the A-Frequent condition chose Category A more than in the Even condition, whereas those in the B-Frequent condition chose Category A less than in the Even condition. The results from the one-way Bayesian ANOVA showed that the data strongly supported the main effect of frequency condition ($BF_{10} = 4.41 \times 10^{10}$). Post hoc comparisons revealed a higher proportion of Category A choices in the A-Frequent condition than in the Even condition ($BF_{10} = 11{,}246$) and the B-Frequent condition ($BF_{10} = 6.21 \times 10^{10}$). Furthermore, the data provided positive evidence that the proportion of Category A choices was higher in the Even condition than in the B-Frequent condition ($BF_{10} = 6.46$).

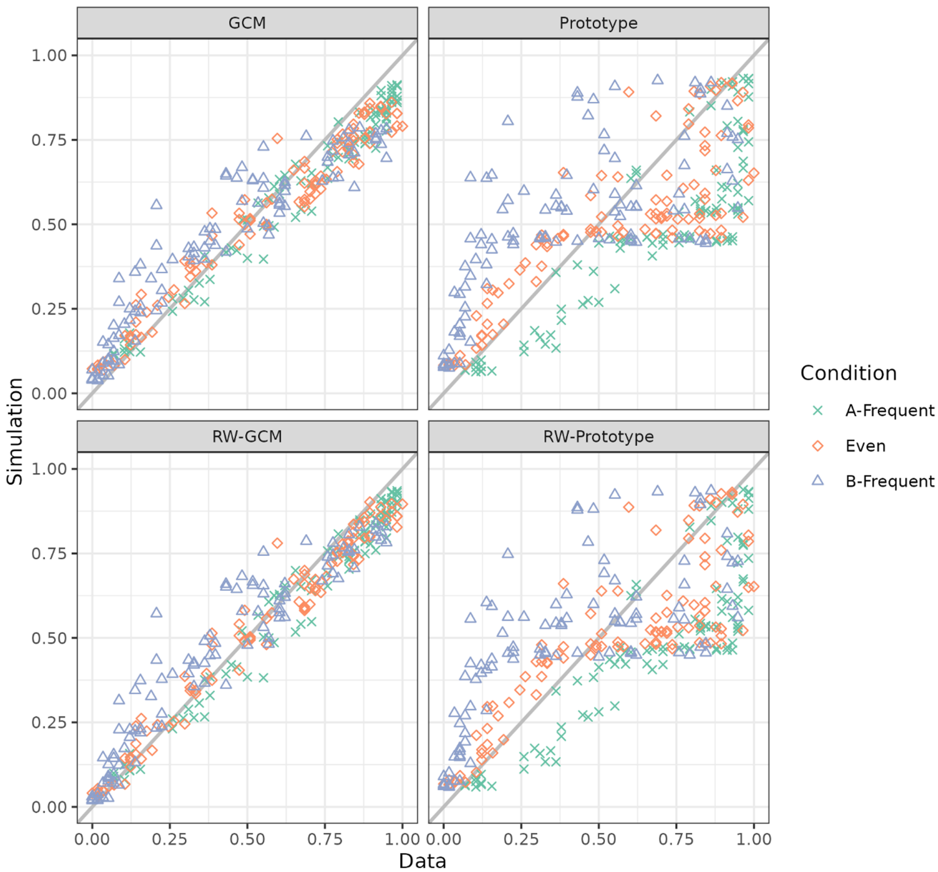
Model-Based Analyses

The mean parameter estimates are summarized in Supplemental Table S2. Table 3 shows the mean BIC, ΔBIC, and Bayes factors across participants for each model for each condition. The Bayes factor comparisons provided strong evidence supporting the better fit of RW-GCM over the other models. We next compared the best-fitting parameters for the RW-GCM among three conditions. No evidence supported a condition difference in the sensitivity (c; $BF_{10} = 0.12$), attention weight to the length dimension (W; $BF_{10} = 0.52$), or decay rate ($1 - \alpha$; $BF_{10} = 0.08$) parameters.

The post hoc simulations revealed that the RW-GCM exhibited the lowest MSD scores of all the models under all conditions. As

Figure 6

Scatterplots of the Simulated Versus Observed Proportion of Category A Choices for Each Stimulus in the Transfer Phase of Experiment 1a



Note. The data are presented separately for each frequency condition. The gray line, with a slope of 1, represents the absolute line of equality. RW = recency-weighted; GCM = generalized context model. See the online article for the color version of this figure.

observed in Experiment 1a, the baseline GCM and the RW-GCM most accurately reproduced the data pattern. Supplemental Figure S5A shows the proportion of Category A choices in the training phase. As predicted, the patterns are extremely similar to those shown in Experiment 1a because the identical category structure was used in the training phase. Supplemental Figure S5B shows the results from the post hoc simulations in the transfer phase. The baseline GCM and the RW-GCM again reproduced the frequency effect on the proportion of Category A choices in the transfer phase. In contrast, the Prototype variants failed to reproduce the pattern of the frequency effect. Supplemental Figure S6 shows the scatterplot detailing the correlations between simulated proportions and observed proportions among three conditions of category frequency in the transfer phase. Like the findings from Experiment 1a (Figure 6), compared to the baseline GCM and RW-GCM, the Prototype models are less accurate in predicting the observed proportions of Category A decisions.

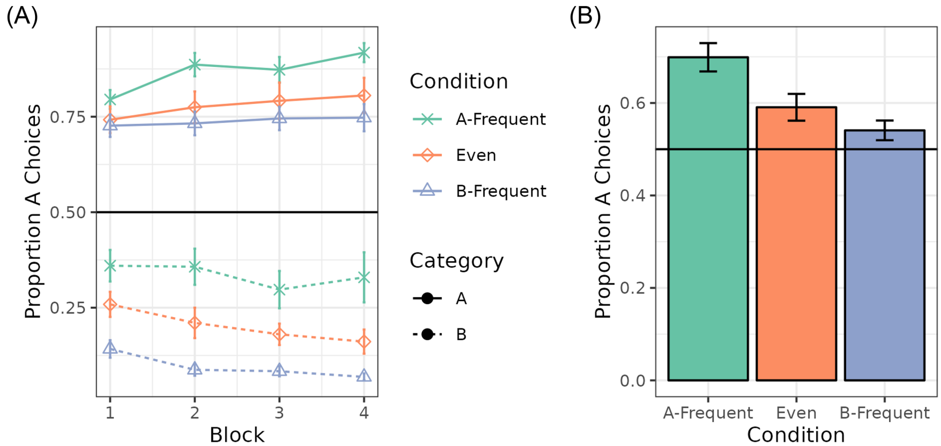
Discussion

Experiment 1 used an II category structure to test for frequency effects and to compare four category learning models in their ability to account for such effects. We found strong frequency effects in both the training and transfer phases. In the training phase, frequency effects were observed for stimuli from both categories. For the transfer phase, there was a bias toward the most-frequent category for both the trained and untrained regions of the stimulus space. This was the case for both sets of transfer stimuli used in Experiments 1a and 1b.

The baseline GCM and RW-GCM, characterized by similarity-accumulating properties, provided a better explanation of the empirical data than the Prototype models. This finding aligns with the assumption from Nosofsky's GCM that the process of categorization is based on accumulating or summing up the similarity between a given stimulus and all previously encountered examples of each category (Nosofsky, 1984, 1986, 1988b). Notably, the RW-GCM outperformed the baseline GCM in both model fitting and post hoc simulation, thereby validating its assumption that more recently seen exemplars receive greater weight when computing the summed similarity to each category (McKinley & Nosofsky, 1995). The Prototype models were the least accurate in accounting for the observed data. Because the Prototype model's similarity metric is derived from the comparison of the novel item and the prototypes for

Figure 7

Mean Proportion of Category A Choices During Training and Transfer Phases Across Three Conditions in Experiment 1b



Note. (A) The mean proportion of Category A choices during the training phase is shown for exemplars from two categories (Category A: solid line; Category B: dashed line), averaged for each 100-trial block across three conditions. (B) The mean proportion of Category A choices during the transfer phase is shown for each condition. The black horizontal line represents a proportion of 0.5. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

each categorization trial, the model does not accumulate similarity representations. For this reason, the Prototype variants cannot account for frequency effects in the present study.

In Experiments 2 and 3, we sought to examine whether the effect of frequency would also be seen in "rule-based" category learning tasks. The category structures used in Experiment 1 are not considered rule-based because the bound separating the two categories is oblique to the perceptual dimensions (Ashby et al., 1998). This makes it difficult to use rules such as "if the line is long and also steep, then it's in Category A. Otherwise, it's in Category B" or "short lines are in Category B, long lines are in Category A." We used rule-based category structures in Experiments 2 and 3 because they were simply a different type of structure with which to examine the effects of frequency on categorization and also because the ability to use verbalizable rules may override some of the effects of frequency. For example, participants might be able to learn an optimal rule to distinguish members of the category, so they may not show as much bias during transfer due to frequency, instead consistently applying the correct rule. Experiment 2 used a conjunctive rule structure where both dimensions were relevant, while Experiment 3 used a simple unidimensional rule along the length dimension.

Table 3

Model Comparison Including Mean BIC, ΔBIC, BF, and MSD in Study 1b

Condition	Model	BIC	ΔBIC	log (BF)	MSD
A-Frequent	GCM	413	28	5.97	0.0077
	Prototype	473	88	19.02	0.0301
	RW-GCM	**385**			**0.0054**
	RW-Prototype	479	94	20.32	0.0322
Even	GCM	449	30	6.49	0.0090
	Prototype	471	52	11.31	0.0209
	RW-GCM	**419**			**0.0054**
	RW-Prototype	470	51	11.10	0.0200
B-Frequent	GCM	367	43	9.40	0.0166
	Prototype	434	110	23.95	0.0433
	RW-GCM	**324**			**0.0083**
	RW-Prototype	436	112	24.38	0.0419

Note. ΔBIC represents the BIC difference between each model to the best model shown in bold. MSD was calculated for all trials in both the training and transfer phases. BF represents the Bayes factor, calculated as exp(ΔBIC/2). The smallest value within each condition is presented in bold. To be brief, we reported the logarithm-transformed value, log (BF), in the table. BIC = Bayesian information criterion; MSD = mean square deviation; RW = recency-weighted; GCM = generalized context model; exp = exponential.

Experiment 2

In Experiment 2, we employed a CJ rule category structure (see the middle panel of Figure 2A) during the training phase to further examine frequency effects in category learning. For the transfer phase, we used the grid structure as applied in Experiment 1b to investigate the impact on participants' generalization of learned categories to novel stimuli.

An appropriate boundary can delineate the CJ category structure within the grid-structured stimulus space. We partitioned the transfer stimuli based on the training category space to determine whether a potential frequency effect occurred in the alternative category space. Some transfer stimuli fell directly on the bound that separated the two categories; these were designated as critical transfer stimuli. Consequently, we could detect the overall frequency effect throughout the entire transfer space, within the opposite category space, and for the critical stimuli along the category bound.

Similar to Experiment 1, we conducted a priori simulations to generate theoretical predictions based on our models of interest, which are reported in Supplemental Figure S7. The simulation results showed that both the baseline GCM and the RW-GCM predicted frequency effects. In contrast, the variants of the Prototype model did not produce the effect.

During the transfer phase, the RW-GCM predicted a larger frequency effect than the baseline GCM. Notably, all four models predicted that the current CJ category structure would cause the agents to be biased toward Category B in the transfer phase. This was expected, as more of the transfer stimuli came from the trained category space for Category B. This aligns with our observations from Experiment 1, where the agents preferred the category that was more dispersed across the stimulus space.

Method

Participants

A total of 254 students at Texas A&M University (170 female participants) with a mean age of 18.57 ($SD = 1.03$; two participants did not report their age) participated in the experiment to partially fulfill a course requirement. Participants who did not achieve at least 50% accuracy in the last 100 trials of the training phase were not included in further analyses to ensure participants achieved the required level of learning. The sample sizes per condition were as follows: 88 in the A-Frequent condition, 79 in the B-Frequent condition (one dropped), and 84 in the Even-Frequent condition (two dropped).

Design

Table 4 indicates category distribution parameters for both categories for the training phase. Stimuli were generated using the same category frequency as in the prior experiments. In the transfer phase, we utilized the identical grid structure employed in Experiment 1b.

Procedure

The task procedure in Experiment 2 was identical to the procedure described in the General Method section.

Results

Behavioral Analyses

As we observed in Experiment 1, the results indicated a successful manipulation of category frequency during training (Figure 8A).

Table 4

Category Distribution Parameters for the Length and Orientation Dimensions in the Conjunctive Category Structure Used in Experiment 2

	Parameter			
Category	μ_l	μ_o	σ_l	σ_o
A	100	200	30	30
B	100	100	30	30
B	200	100	30	30
B	200	200	30	30

Note. Dimensions are in arbitrary units. Please refer to Figure 2B for the scatterplot. The subscripts "l" and "o" denote length and orientation, respectively.

The 3 (Condition) × 4 (Block) Bayesian Repeated Measures ANOVAs were performed to test the frequency effects. The results revealed that the data strongly favored the model that included the interaction between Block and Condition for Category A stimuli ($BF_{10} = 1.89 \times 10^{42}$) and for Category B stimuli ($BF_{10} = 1.68 \times 10^{43}$).

The frequency effect was detected in the transfer phase (Figure 8B). As observed in Experiment 1, relative to the Even condition, participants selected Category A more often in the A-Frequent condition and selected A less in the B-Frequent condition. A one-way Bayesian ANOVA showed that the data strongly supported the model that included the main effect of condition compared with the null model ($BF_{10} = 5.63 \times 10^{33}$). Post hoc comparisons revealed a higher proportion of Category A choices in the A-Frequent condition compared to both the Even condition ($BF_{10} = 2.76 \times 10^{10}$) and the B-Frequent condition ($BF_{10} = 6.3 \times 10^{26}$) and a higher proportion of Category A choices in the Even condition compared to the B-Frequent condition ($BF_{10} = 3.72 \times 10^{12}$).

We next examined frequency effects for stimuli located in various areas of the transfer stimulus space. Figure 8C illustrates the three transfer categories. Two regions are mapped according to the two spaces of training categories. The "critical" region contains the stimuli located along the boundary between the two categories. The proportion of Category A choices across the three distinct regions of the stimulus space is displayed in Figure 8D. As observed in Figure 8B for the whole stimulus space, this frequency effect was notably discernible within each region of the transfer space. A 3 (Space) × 3 (Condition) Bayesian Repeated Measures ANOVA was performed and revealed that the data strongly support the model including the Space × Condition interaction ($BF_{10} = 2.64 \times 10^{190}$).

Three simple main effects were analyzed for each training category separately. The data strongly supported the main effect of frequency condition in the three training categories. For stimuli in the Category A trained space ($BF_{10} = 1.02 \times 10^{13}$), post hoc analyses revealed that the participants in the A-Frequent condition made more Category A choices than participants in the Even condition ($BF_{10} = 104$), and participants in the Even condition had more A selections than participants in the B-Frequent condition ($BF_{10} = 26{,}112$). For stimuli in the B-trained space ($BF_{10} = 1.9 \times 10^{19}$), post hoc comparisons indicated that there were more Category A choices in the A-Frequent condition than in the Even condition ($BF_{10} = 4.29 \times 10^{6}$) and in the Even condition than in the B-Frequent condition ($BF_{10} = 108{,}613$). For the critical transfer stimuli ($BF_{10} = 1.52 \times 10^{36}$), post hoc comparisons revealed more Category A choices in the A-Frequent condition than the Even condition ($BF_{10} = 2.2 \times 10^{11}$) and in the Even condition than the B-Frequent condition ($BF_{10} = 2.01 \times 10^{9}$).

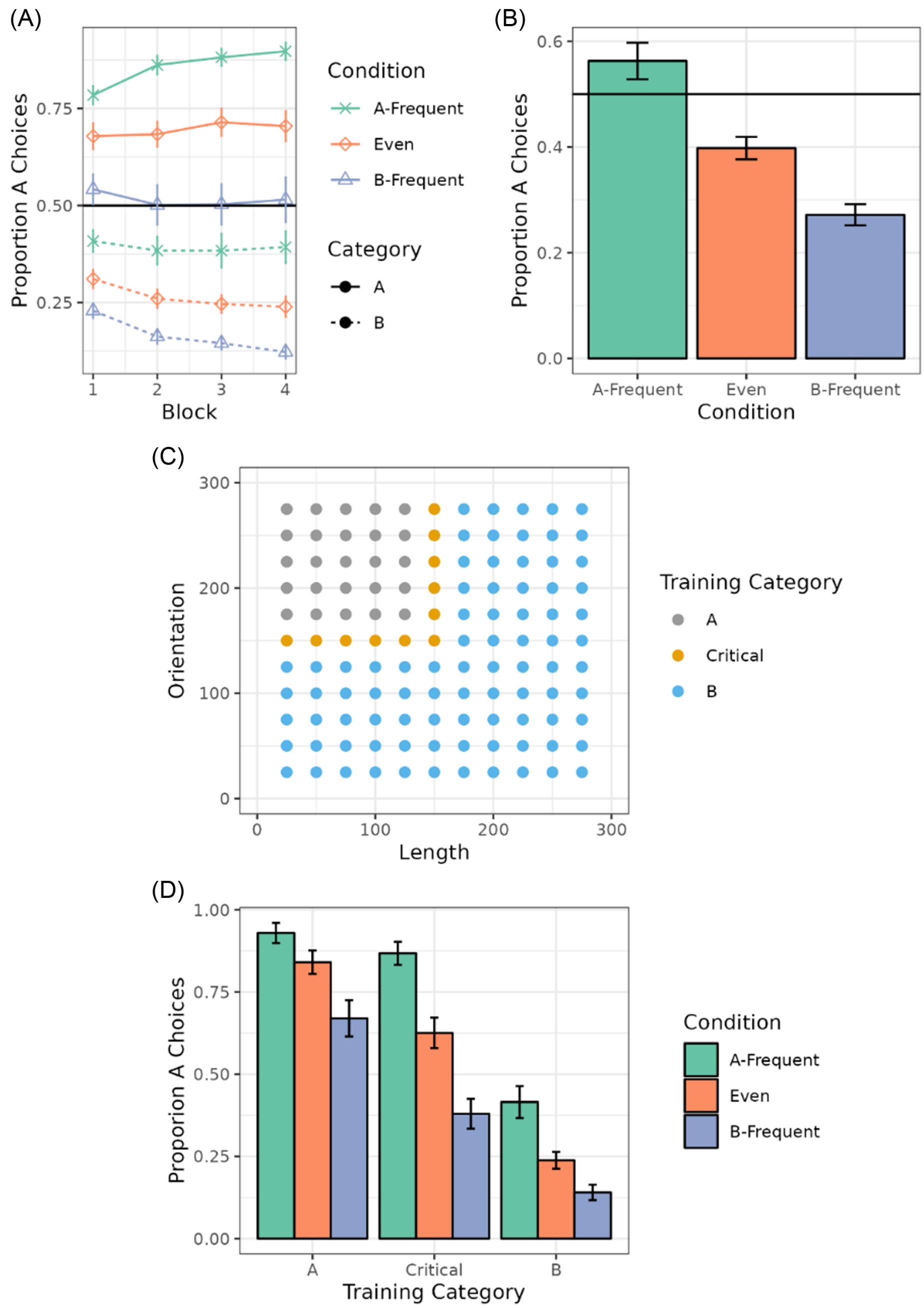
Model-Based Analyses

We fit the same four models used in Experiment 1. The estimated parameters are indicated in Supplemental Table S3. Table 5 reveals the mean BIC, ΔBIC, and Bayes factors across participants for each model for each condition. Again, the RW-GCM best fits the data in all three conditions. Additionally, a Bayesian one-way ANOVA revealed no significant evidence of differences across conditions in the sensitivity (c; $BF_{10} = 0.24$), attention weight to the length dimension (W; $BF_{10} = 0.04$), or the decay rate ($1 - \alpha$; $BF_{10} = 0.11$) parameters within the RW-GCM.

The post hoc simulation results are presented in the Supplemental Materials. Again, as we observed in Experiments 1a and 1b, the

Figure 8

Mean Proportion of Category A Choices During Training and Transfer Phases Across Three Conditions in Experiment 2



Note. (A) The mean proportion of Category A choices during the training phase is shown for exemplars from each category (Category A: solid line; Category B: dashed line), averaged across each 100-trial block across three conditions. (B) The mean proportion of Category A choices during the transfer phase is shown for each condition. (C) The diagram shows the mapping of training categories onto the transfer space. Gray dots represent Category A, with 25 trials. Blue dots represent Category B, with 85 trials. Orange dots indicate critical stimuli, with 11 trials. (D) The mean proportion of Category A choices is shown across three transfer spaces. The black horizontal line represents a proportion of 0.5. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

Table 5

Model Comparison Including Mean BIC, ΔBIC, BF, and MSD in Experiment 2

Condition	Model	BIC	ΔBIC	log (BF)	MSD
A-Frequent	GCM	444	18	3.97	0.0061
	Prototype	483	58	12.52	0.0188
	RW-GCM	**425**			**0.0055**
	RW-Prototype	477	52	11.21	0.0166
Even	GCM	520	12	2.60	0.0098
	Prototype	517	9	1.87	0.0104
	RW-GCM	**508**			**0.0092**
	RW-Prototype	515	7	1.43	0.0111
B-Frequent	GCM	482	11	2.45	0.0138
	Prototype	562	91	19.82	0.0436
	RW-GCM	**471**			**0.0133**
	RW-Prototype	570	99	21.55	0.0458

Note. ΔBIC represents the BIC difference between each model to the best model shown in bold. MSD was calculated for all trials in both the training and transfer phases. BF represents the Bayes factor, calculated as exp(ΔBIC/2). The smallest value within each condition is presented in bold. To be brief, we reported the logarithm-transformed value, log (BF), in the table. BIC = Bayesian information criterion; MSD = mean square deviation; RW = recency-weighted; GCM = generalized context model; exp = exponential.

RW-GCM demonstrated the lowest MSD scores of all the models across all situations (see Table 5). The baseline GCM and the RW-GCM exhibited higher predictive accuracy for the empirical data across the various conditions than the two Prototype models (see Supplemental Figures S9, S10).

Discussion

As in Experiment 1, our manipulation of category frequency influenced participants' categorization tendencies during the training and transfer phases. This effect was observed both at the boundary between the two categories and within the space belonging to the opposing category. The model-based analyses revealed that the GCM variants, which assume similarity accumulation, exhibited superior performance in model fitting and replicating the observed categorization patterns than the Prototype models. The RW-GCM offered the most accurate explanation of the observed data, displaying the lowest BIC values in three conditions and lower MSD scores in nearly all conditions when compared to the baseline GCM. The lower MSD scores from the post hoc simulations for the RW-GCM suggest that its assumption of a recency bias may allow it to better account for frequency effects.

Experiment 3

In Experiment 3, we utilized a UD category structure (right panel of Figure 2A) during the training phase to examine the potential impact of frequency on categorization involving a simple, unidimensional rule. Compared with the previously adopted category structures in Experiments 1 and 2, the UD structure is much simpler, and the rule is likely easier to learn than the conjunctive rule. We thus examined whether frequency effects would still be observed when a relatively simple verbalizable rule could distinguish the categories. Additionally, since the two UD categories utilized only a small portion of the space, we could investigate whether the frequency effect was evident in stimuli within the untrained regions of the stimulus space.

As we observed from a priori simulations in Experiments 1 and 2, the two variants of the GCM with a similarity-accumulating attribute (baseline GCM and RW-GCM) again predicted frequency effects in the training and transfer phases. As noted in Experiments 1 and 2, the Prototype variants did not predict frequency effects. The simulation results are reported in detail in Supplemental Figure S11. These consistent predictions across different experiments reinforce the robustness of the GCM and RW-GCM models in capturing the frequency effects observed in our experiments.

Method

Participants

Participants were 243 students at Texas A&M University (154 female participants, three preferred not to respond) with a mean age of 18.56 (SD = 0.89; four participants did not report their age) who participated in experiments to partially fulfill a course requirement. Participants who did not achieve at least 50% accuracy in the last 100 trials of the training phase were not included in further analyses. The sample sizes per condition were as follows: 80 in the A-Frequent condition (two dropped), 76 in the B-Frequent condition (two dropped), and 67 in the Even-Frequent condition (16 dropped).

Design

Table 6 indicates category distribution parameters for both categories in the UD structure. In the transfer phase, Experiment 3 employed an identical grid structure to that used in Experiments 1b and 2.

Procedure

The task procedure was identical to those described in the General Method section.

Results

Behavioral Analyses

The results from the training phase of Experiment 3 are indicated in Figure 9A. We again observed the frequency effects on the proportion of Category A choices for each category. The 3 (Condition) × 4 (Block) Bayesian Repeated Measures ANOVAs revealed strong evidence supporting the model that included an interaction between

Table 6

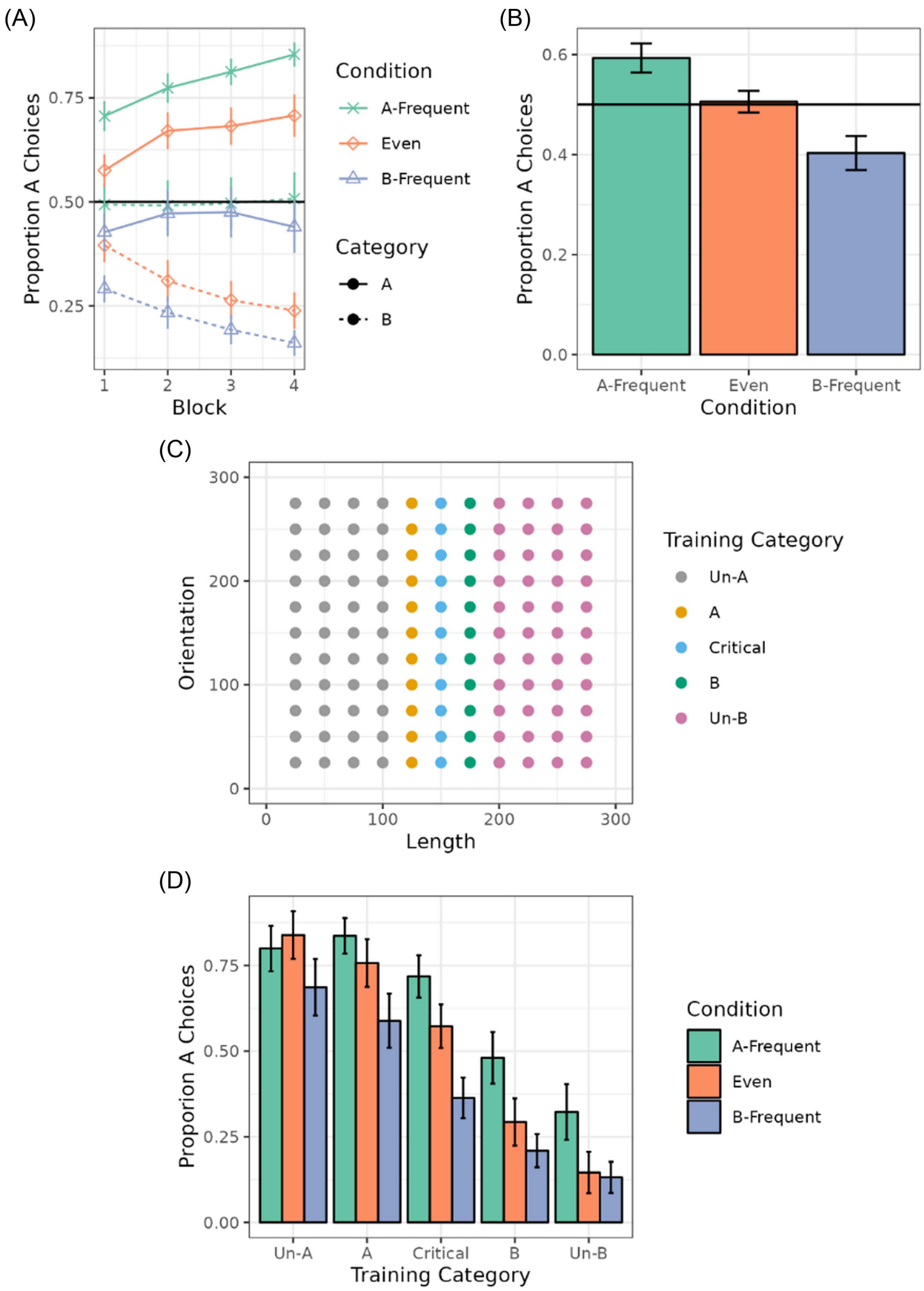
Category Distribution Parameters for the Length and Orientation Dimensions in the Unidimensional Category Structure Used in Experiment 3

Category	Parameter			
	μ_l	μ_o	σ_l	σ_o
A	140	150	5	55
B	160	150	5	55

Note. Dimensions are in arbitrary units. The subscripts "l" and "o" denote length and orientation, respectively.

Figure 9

Mean Categorization Accuracy Rates for Two Categories Across Four Blocks in Three Conditions During the Training Phase in Experiment 3



Note. (A) The mean proportion of Category A choices during the training phase is shown for exemplars from two categories (Category A: solid line; Category B: dashed line), averaged for each 100-trial block across three conditions. (B) The mean proportion of Category A choices during the transfer phase is shown for each condition. (C) The diagram shows the mapping of training categories onto the transfer space. Gray dots represent the Untrained-A side (Un-A: 44 trials); orange dots represent training Category A (A: 11 trials); blue dots represent critical stimuli (11 trials); green dots represent training Category B (B: 11 trials); and purple dots represent the Untrained-B side (Un-B: 44 trials). (D) The mean proportion of Category A choices is shown across all transfer spaces. The black horizontal line represents a proportion of 0.5. Error bars show 95% confidence intervals. See the online article for the color version of this figure.

Block and Condition for Category A stimuli ($BF_{10} = 2.72 \times 10^{45}$) and Category B stimuli ($BF_{10} = 9.94 \times 10^{32}$), indicating a varying frequency effect across blocks between conditions.

We again observed a frequency effect in the transfer phase of Experiment 3 (Figure 9B). The results of one-way Bayesian ANOVA show that the data most strongly support the model that includes the main effect of condition ($BF_{10} = 7.04 \times 10^{13}$). Post hoc analyses highlighted a notably higher proportion of Category A selections in the A-Frequent condition than the Even condition (BF_{10} = 2,223) and the B-Frequent condition ($BF_{10} = 3.32 \times 10^{11}$). Additionally, there was a greater proportion of Category A choices for participants in the Even than in the B-Frequent condition (BF_{10} = 6,775).

We also investigated the frequency effect on different training stimulus spaces. Figure 9C illustrates the five transfer categories. We mapped these categories according to the stimulus space for the trained categories. Stimuli situated in two untrained spaces adjacent to Categories A and B were assigned to two untrained categories: Untrained-A and Untrained-B, respectively. Additionally, two transfer categories were mapped based on the spaces in Categories A and B during the training phase. Finally, stimuli positioned in the middle of the boundary space between the two training categories were designated as critical stimuli.

Figure 9D illustrates the proportion of Category A choices for each training space among three conditions in the transfer phase. The pattern of a frequency effect can be identified in most spaces. A 5 (Space) × 3 (Condition) Bayesian Repeated Measures ANOVA indicated that the data strongly support the model that includes the Space × Condition interaction ($BF_{10} = 1.43 \times 10^{125}$).

The simple main effects were performed for each space. BF results show strong support for the main effect of condition for the trained Category A stimuli (BF_{10} = 14,117); post hoc tests within the trained Category A space showed that participants were not more likely to choose Category A in the A-Frequent condition than in the Even condition ($BF_{10} = 0.85$), but that participants were more likely to choose Category A in the Even condition than in the B-Frequent condition ($BF_{10} = 16$). In the trained Category B space (BF_{10} = 421,498), post hoc tests indicated that participants were more likely to choose Category A in the A-Frequent condition than in the Even condition ($BF_{10} = 57$), but there was only weak evidence that there were more A choices in the Even condition than in the B-Frequent condition ($BF_{10} = 1.1$).

For the critical stimuli ($BF_{10} = 9.81 \times 10^{10}$), post hoc analyses showed positive evidence for a higher proportion of Category A choices in the A-Frequent condition than in the Even condition ($BF_{10} = 20$). Strong evidence also supported a higher proportion of Category A choices in the Even condition than in the B-Frequent condition (BF_{10} = 3,853).

In the Untrained-A space, weak evidence supported the main effect of the condition of frequency ($BF_{10} = 2.6$), and positive evidence supported a higher proportion in the Even condition compared to the B-Frequent condition ($BF_{10} = 6$). However, strong evidence supported the main effect of frequency in the Untrained-B space ($BF_{10} = 653$). Post hoc tests indicated that the A-Frequent condition surpassed the Even condition in the proportion of Category A choices ($BF_{10} = 28$). Overall, the pattern of results suggests that the frequency manipulation led to the preference for the high-frequency category when the stimulus was from the other category's trained space.

Model-Based Analyses

The best-fitting parameters are shown in Supplemental Table S4. Table 7 reveals the mean BIC, ΔBIC, and Bayes factors across participants for each model for each condition. The RW-GCM exhibited the lowest BIC in both the A-Frequent and B-Frequent conditions. Bayes factor results revealed strong evidence supporting a superior fit of the RW-GCM compared to the other models in these two conditions. Unexpectedly, the RW-Prototype displayed the lowest BIC value in the Even condition, diverging from findings in Experiments 1 and 2.

We also examined the parameter estimations for the RW-GCM across the three conditions, as we reported in previous experiments. Bayesian one-way ANOVA revealed no evidence supporting a condition difference in the sensitivity (c; $BF_{10} = 0.21$), attention weight to the length dimension (W; $BF_{10} = 0.24$), or decay rate ($1 - \alpha$; $BF_{10} = 0.67$) parameters.

Table 7 indicates MSD scores from each model in the three conditions. The findings showed that the RW-GCM had the lowest MSD scores compared to the other models in the A-Frequent and the B-Frequent conditions. In the Even condition, the RW-Prototype exhibited lower MSD scores than the RW-GCM, although their scores were very similar, differing by less than .001.

The post hoc simulation results are presented in the Supplemental Materials. Supplemental Figure S13 shows that the baseline GCM and the RW-GCM reproduced the pattern of the frequency effect to the observed patterns in the data in the training phase (Supplemental Figure S13A) and the transfer phase (Supplemental Figure S13B). On the contrary, the Prototype variants could not generate this pattern and predicted equal selections of each category in the three conditions. Supplemental Figure S14 provides a scatterplot illustrating the relationships between simulated and observed proportions of Category A choices in the transfer phase across the three conditions. As observed in previous experiments, the baseline GCM and the RW-GCM predicted the empirical data more accurately

Table 7

Model Comparison Including Mean BIC, ΔBIC, BF, and MSD in Experiment 3

Condition	Model	BIC	ΔBIC	log (BF)	MSD
A-Frequent	GCM	562	35	7.61	0.0090
	Prototype	575	48	10.39	0.0202
	RW-GCM	**527**			**0.0051**
	RW-Prototype	570	43	9.30	0.0207
Even	GCM	605	46	10.08	0.0162
	Prototype	576	17	3.69	0.0074
	RW-GCM	572	13	2.80	0.0074
	RW-Prototype	559			**0.0068**
B-Frequent	GCM	570	27	5.96	0.0121
	Prototype	598	56	12.14	0.0250
	RW-GCM	**542**			**0.0075**
	RW-Prototype	592	50	10.84	0.0253

Note. ΔBIC represents the BIC difference between each model to the best model shown in bold. MSD was calculated for all trials in both the training and transfer phases. BF represents the Bayes factor, calculated as exp(ΔBIC/2). The smallest value within each condition is presented in bold. To be brief, we reported the logarithm-transformed value, log (BF), in the table. BIC = Bayesian information criterion; MSD = mean square deviation; RW = recency-weighted; GCM = generalized context model; exp = exponential.

than the Prototype models in the conditions where frequency was manipulated.

Discussion

As in Experiments 1 and 2, we found strong effects of frequency, even when participants categorized novel stimuli in the transfer phase, including the stimuli from the category space of the alternative category. The two variants of the GCM (baseline GCM and RW-GCM) predicted a strong preference toward the high-frequency category, due to their similarity-accumulating properties. In the transfer phase, both variants of the GCM could reproduce the pattern of the observed data. In contrast, the Prototype models predicted a 50% proportion of Category A choices in all three conditions, suggesting that the Prototype variants failed to predict participants' categorization decisions in the transfer phase. The RW-GCM showed lower MSD scores in all three conditions than the baseline GCM, suggesting that the additional assumption of a bias toward more recent exemplar can better capture frequency effects, although, interestingly, the RW-Prototype provided the best fit to the data in the Even condition.

General Discussion

We conducted four independent experiments to explore the assumptions of prototype and exemplar models in human category learning, focusing specifically on the influence of category frequency. Across experiments, different category structures (II, CJ, and UD structure; see Figure 2A) were employed in the training phase. During the transfer phase, participants classified novel stimuli generated from a broader area encompassing most of the stimulus space utilized in this study. Across all experiments, significant frequency effects were observed where participants classified novel items into the category that appeared more frequently than the other. Experiment 1a revealed that this effect extended to stimuli from the untrained space. Experiment 2 indicated the effect's presence within items from the opposing (low-frequency) category, while Experiment 3 identified the frequency effect within stimuli from the untrained region of the space.

To examine how well variants of the exemplar and prototype models could account for frequency effects in human category learning, we evaluated both baseline models and recency-weighted variants of the GCM (RW-GCM) and the Prototype model (RW-Prototype) to investigate the computational mechanisms underlying categorization when one category was presented more frequently than another. The RW-GCM employs the Decay learning rule (Erev & Roth, 1998), which accumulates recency-weighted similarity for exemplars as it compares the current stimulus to each exemplar from each category. As the model computes similarity to each exemplar, from the first seen in each category to the most recently seen, summed similarity decays so that the model gives greater weight to the similarity to the most recent exemplars from each category. On the other hand, the RW-Prototype applies the Delta rule (Rescorla & Wagner, 1972; Widrow & Hoff, 1960) to update the category prototypes by computing recency-weighted averages for the stimulus dimensions of each category. The results demonstrated that the RW-GCM provided the best fit to the data and successfully reproduced observed classification patterns across all experiments. In contrast, prototype models failed to account for the frequency effects, while the baseline GCM explained frequency effects but not recency effects.

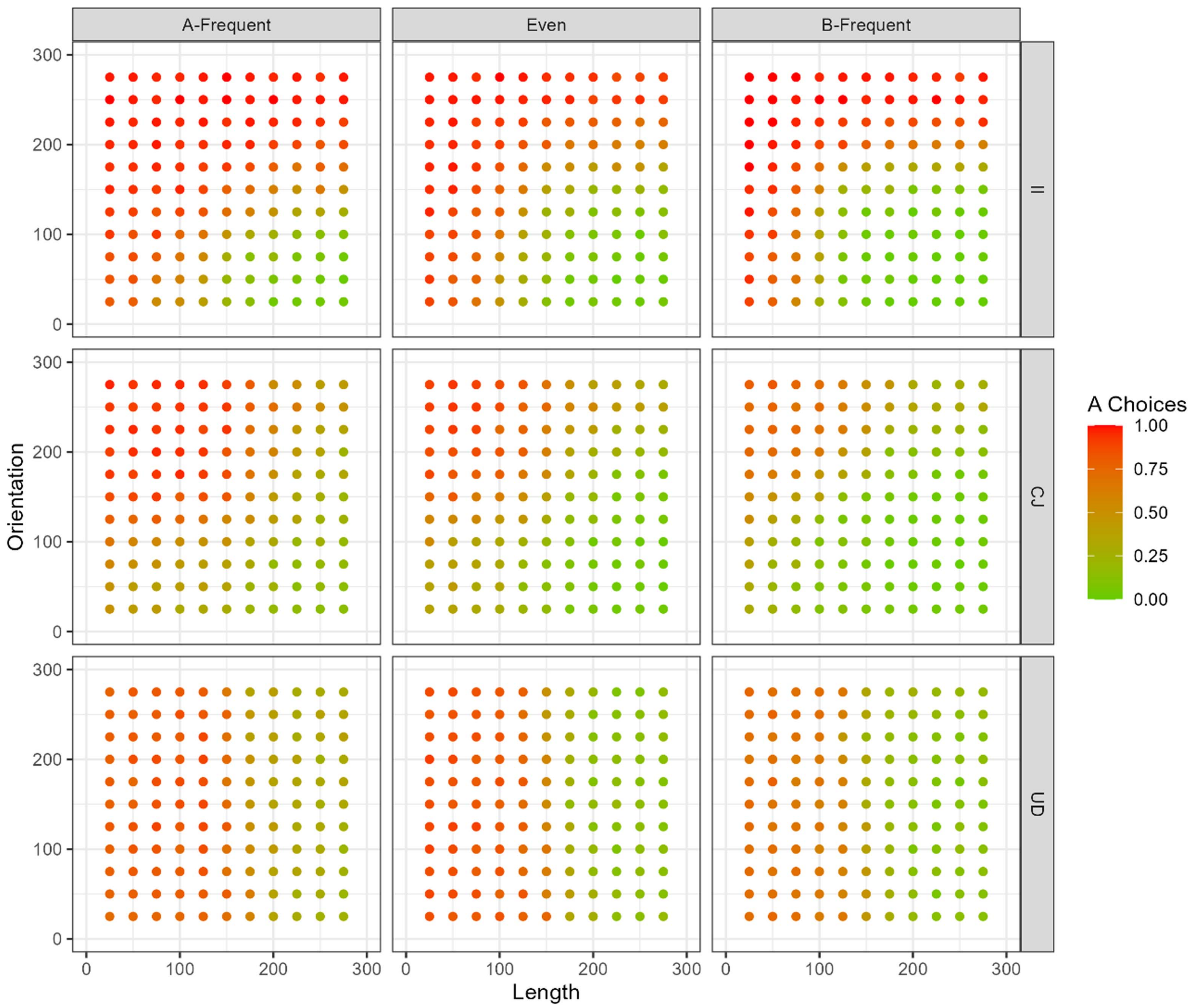
Our findings resonate with Nosofsky's (1988b) work, which demonstrated participants' higher "goodness-of-example" judgments for frequently encountered exemplars and the generalization of these preferences to neighboring items in the stimulus space. Similarly, we observed a preference for high-frequency categories during the transfer phase, with the frequency effect extending to broader regions of the stimulus space, even those unexposed during training. Figure 10 illustrates that category space expansion is caused by the frequency effect in the transfer phase across three experiments. The A-Frequent condition displays a more substantial red area compared to the Even condition. Conversely, the B-Frequent condition reveals a more pronounced green area relative to the Even condition, thereby evidencing the frequency effect of opting for the more prevalent training category in the transfer phase. It depicts how the high-frequency category extends the categorical boundary beyond its trained region.

Frequency Effects in Other Paradigms

The frequency effect, evidenced by accumulated representation, was similarly observed in earlier studies such as those by Estes (1976a, 1976b). In Estes's observation learning task, participants were asked to observe a series of trials where pairs of stimuli (e.g., political candidates) were shown alongside winning or losing outcomes. Notably, participants were observers only and made no selections. Then, they predicted preferences in a hypothetical survey between an alternative stimulus in a pair from the different combinations of stimuli. Estes found that more frequently presented stimuli were favored over less-frequent ones, irrespective of a higher reward probability for the latter. A similar frequency effect was also reported by Don et al. (2019), where participants undertook a reward learning task with a similar frequency structure. Despite having a lower average reward probability than less-frequent alternatives, options encountered more frequently were preferred. Additionally, model-based analysis from the study revealed that a model incorporating a Decay rule better accounted for the frequency effect than one using a Delta rule, which aligns with our modeling results. Both findings suggest that people are sensitive to event frequency, and the options or actions might be valued based on their cumulative reward, rather than average reward.

The decision from experience literature has revealed that participants tend to select the options that produce more frequent optimized outcomes, even when those options yield lower expected payoffs in the long run (Barron & Erev, 2003; Hertwig et al., 2004; Yechiam & Busemeyer, 2005). This phenomenon often manifests as a tendency to assign lower importance to rare events when making decisions based on experiential knowledge. Recent research suggests that this underweighting of rare events in experiential decision making may stem from a reliance on limited data samples (Erev et al., 2010; Hertwig & Erev, 2009; Rakow & Newell, 2010). As a result, participants may rely on a relatively small set of past exemplars stored in memory to decide their next categorization decision, giving greater weight to the stimuli that occurred recently relative to earlier exemplars. That might explain why the variant of the GCM, which accounts for recency, the RW-GCM, outperformed the baseline GCM.

Figure 10
Illustration of the Proportion of Category A Choices for Each Stimulus in the Transfer Phase Across Three Training Category Structures



Note. The color bar represents the proportion of times Category A was selected by participants. A shift toward red indicates a higher selection rate of Category A, while leaning toward green signifies a greater selection rate of Category B. II = information-integration; CJ = conjunctive; UD = unidimensional. See the online article for the color version of this figure.

Recency Weighting

The concept of recency weighting implies that both memory traces and exemplar representations decay over time, and there is support for this idea from several studies in the memory literature. For instance, Ebbinghaus's renowned forgetting curve illustrates how individuals forget information over extended periods (see Murre & Dros, 2015). This pattern of representation degradation over time is not exclusive to long-term memory; it is also observed in short-term memory (Atkinson & Shiffrin, 1968; Peterson & Peterson, 1959), sensory (iconic) memory (Sperling, 1960), and visual working memory (Broadbent & Broadbent, 1981). The temporal issue may not solely cause the fading of representations. Interference from preceding or subsequent information can also precipitate this decay. For example, retroactive interference happens when a new stimulus hinders the retrieval of previously stored representations (e.g., Baddeley & Dale, 1966), while proactive interference occurs when memory traces disrupt the retrieval of new information (e.g., Wickens et al., 1963). In our category learning tasks, participants repeatedly encountered similar stimuli, a process during which retroactive and proactive interferences were likely to occur. Apart from temporal decay, interference from new items undoubtedly contributes to the observed recency effect. However, it is challenging to disentangle the effects of temporal decay and interference solely based on the recency parameter in our best-fitting model, the RW-GCM. Further research is required to address this question adequately.

Exemplar Versus Prototype Models

Can prototype-based models take account of the frequency effects identified in this study? As we introduced before, classic research using the dot-pattern prototype-distortion paradigm showed participants often displayed higher classification accuracy for nontrained prototypes than new distortions, and sometimes even trained distortions, in the transfer phase—evidencing the prototype-enhancement effect foundational to prototype abstraction (Donald et al., 1973; Posner & Keele, 1968, 1970). Homa et al. (1981, 1991) and Homa (1984) reported an increased prototype-enhancement effect with larger category sizes, and an old–new similarity effect diminished with increased category size. While these studies imply a category frequency effect on accurate classifications in the transfer phase, they did not show a frequency effect on the tendency to select high-frequency categories when classifying new or prototype patterns in the transfer phase (e.g., Homa et al., 1981, Table 2), unlike what we observed in our experiments. Several procedural and methodological differences exist between the dot-pattern paradigm and our category learning task. For instance, their category sizes were relatively small (e.g., 3, 6, 9 or 5, 10, 20 per category) compared with ours (134, 266). Also, their training exemplars (old distortions) were repeated many times in each block during the learning phase, leading to high accuracy rates by the end of learning—a prerequisite for progression in their paradigm. This restricted number of exemplars and high repetition could explain the emergence of the prototype-enhancement effect and the absence of a high-frequency category selection tendency in the transfer phase.

While the exemplar model can account for the classical category size effects observed in dot-pattern paradigm studies (Nosofsky, 1988a; Shin & Nosofsky, 1992), prototype-based theories appear to struggle to explain the frequency effects observed in our category learning tasks. The GCMs, which include a similarity-accumulating property, captured and predicted participants' tendencies to categorize novel objects into high-frequency categories because these categories accumulated relatively more summed similarity to previously experienced exemplars. In contrast, prototype models (Donald et al., 1973; Reed, 1972; Smith & Minda, 1998) propose the abstraction of a singular, central-tendency prototype from learned exemplars, irrespective of their number, meaning there's potentially just one prototype representing each category. Consequently, novel objects are classified based on their similarity to this prototype, independent of the amount of the exemplars belonging to any one category. Thus, it appears that the prototype model is unable to adequately explain our findings, without some major modification to the model.

The Strong Effect of Base Rates

Our results elucidate the base rate's significant role in decision making, particularly highlighting the cognitive tendency that can arise from category frequency. One of the most salient applications of this principle is the availability heuristic, introduced by Tversky and Kahneman (1973). This heuristic refers to a common mental shortcut wherein individuals assess the frequency or probability of an event based on the ease with which examples or occurrences can be recalled. The word frequency effect offers a compelling demonstration of this phenomenon in the realm of language processing. Research indicates that words that appear more frequently in a language (high-frequency words) are recognized and processed more rapidly than those that are less common (low-frequency words), as detected in various reading tasks (e.g., Monsell et al., 1989).

In making categorization decisions, people may lean toward selecting more frequently encountered categories, a tendency that can be attributed to the category's heightened availability in memory. This inclination facilitates rapidly generalizing unfamiliar objects into categories and engages the intuitive, quick-response mechanism. For instance, upon encountering an unfamiliar object with sharp edges, people instinctively relate it to previously encountered objects known to be sharp—and often dangerous. Due to their frequent association with risk or harm, these familiar objects are readily available in our memory, allowing for a quick comparison. Consequently, the unfamiliar object is rapidly categorized as "dangerous," triggering a state of heightened alertness during interaction. This process highlights the essential role of the availability heuristic in our cognitive toolkit, serving as a protective mechanism by enabling us to identify potential threats quickly in our environment.

Limitations

We identified frequency effects and presented evidence that the RW-GCM provides the best account for these effects in our category learning task. Nevertheless, some limitations to our models and results should be addressed in future work. The current variants of the GCM are applied only to situations in which the stimuli can be represented as points in a multidimensional space. Our results hinge on the assumption that the foundational dimensions remain constant during the training phase, and the model does not extend to scenarios where participants might create new dimensions of representation or recode the stimuli as they learn. Future studies might consider employing categories with psychologically integral stimuli, such as variations in brightness and saturation, to explore whether the frequency effects persist with these stimulus types.

Despite our task's stimuli being constructed from a two-dimensional space, our approach did not leverage multidimensional scaling solutions to map the stimulus space into a psychological one (e.g., Nosofsky, 1986), a significant presumption within the GCM framework. We employed a mass number of stimuli in our category learning task; thus, collecting required similarity judgments for the multidimensional scaling analysis across all pairs of stimuli was impractical. Consequently, our computations of distances and similarities were based solely on the physical attributes of the stimuli. However, the stimuli, varying in line length and orientation, were psychologically distinguishable in our task. We operated under the assumption that the physical stimulus space closely mirrors the perceptual space used by participants when making categorization judgments.

It is also important to note that frequency effects have been found in many simpler tasks such as probability matching (Estes, 1976a), and it is possible that participants were simply matching their responses to the frequency of feedback. The models we used did not include a response bias parameter, but the addition of a response bias parameter would likely improve the ability of the prototype models to account for frequency effects. However, a constant response bias toward the more frequent category could lead to slightly different predictions than the frequency bias predicted by the summed similarity mechanism of the GCM, which will differ in strength, as more stimuli are observed in the task. Future research could attempt to address this issue by including response bias parameters in these models and designing experiments that can address whether frequency effects are due more to a consistent bias or difference in summed similarity between the two categories.

Future Directions

Future studies might explore variations of the GCM that track how participants shift their attention among different dimensions on a trial-by-trial basis during category learning. The original GCM (Nosofsky, 1986) presupposes that attention weight is static, established as a free parameter for estimation. However, participants may allocate their attention between different dimensions at different stages of learning. For instance, in Experiment 3 of the present study, which utilized a unidimensional category structure, participants may have alternated their attention between two dimensions. As they progressed through the latter blocks of the training phase and began assimilating category knowledge, they may have concentrated more on the length of each line. The GCM's successors, like attention-learning coverage map (Kruschke, 1992) and Supervised and Unsupervised STratified Adaptive Incremental Network (Love et al., 2004), both neural network–based models, incorporate mechanisms that adjust attention weights based on learning from errors postexperience. Future work should also consider whether these network models and other models not examined here can account for the seemingly strong effects of both frequency and recency.

Conclusion

To conclude, our findings highlight the impact of frequency effects on human categorization performance, demonstrating a preference among participants to classify novel items based on the category more frequently presented during the learning stage. This effect was identified across studies that used different category structures. Model-based analyses favored the RW-GCM, which assumed that the representation of similarity for a given category is accumulated with recency weighting. This variant was found to account for the frequency effect in our category learning tasks more effectively than the baseline GCM. Conversely, the Prototype models did not fit as well as the GCM models and failed to replicate the observed frequency effect. Our findings suggest that category frequency significantly influences categorization across both unverbalizable, implicit structures and verbalizable, rule-based structures. The superior fit of the RW-GCM implies decay in category representations over time. Furthermore, our findings support the idea that implementing reinforcement learning rules within the GCM framework can enhance the model's efficacy.

References

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., & Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science*, *32*(8), 1376–1402. https://doi.org/10.1080/03640210802352992

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53. https://doi.org/10.1037/0278-7393.14.1.33

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400. https://doi.org/10.1006/jmps.1993.1023

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). Academic Press.

Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *5*(5), 417–420. https://doi.org/10.1016/S0022-5371(66)80054-3

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233. https://doi.org/10.1002/bdm.443

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

Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effects in visual memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *33*(1), 1–15. https://doi.org/10.1080/14640748108400762

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–215). Lawrence Erlbaum.

Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*, Article 505. https://doi.org/10.3389/fpsyg.2015.00505

Cornwall, A. C., Davis, T., Byrne, K. A., & Worthy, D. A. (2022). Effects of categorical and numerical feedback on category learning. *Cognition*, *225*, Article 105163. https://doi.org/10.1016/j.cognition.2022.105163

Daniel, R., & Pollmann, S. (2010). Comparing the neural basis of monetary reward and cognitive feedback during information-integration category learning. *The Journal of Neuroscience*, *30*(1), 47–55. https://doi.org/10.1523/JNEUROSCI.2205-09.2010

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Don, H. J., Otto, A. R., Cornwall, A. C., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition*, *193*, Article 104042. https://doi.org/10.1016/j.cognition.2019.104042

Don, H. J., & Worthy, D. A. (2022). Frequency effects in action versus value learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1311–1327. https://doi.org/10.1037/xlm0000896

Donald, H., Joseph, C., Don, C., David, G., & Steven, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*(1), 116–122. https://doi.org/10.1037/h0035772

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., Hertwig, R., Stewart, T., West, R., & Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*(1), 15–47. https://doi.org/10.1002/bdm.683

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, *88*(4), 848–881. https://www.jstor.org/stable/117009

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*(2), 94–107. https://doi.org/10.1037/h0058559

Estes, W. K. (1957). Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika*, *22*(2), 113–132. https://doi.org/10.1007/BF02289048

Estes, W. K. (1976a). The cognitive side of probability learning. *Psychological Review*, *83*(1), 37–64. https://doi.org/10.1037/0033-295X.83.1.37

Estes, W. K. (1976b). Some functions of memory in probability learning and choice behavior. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 10, pp. 1–45). Academic Press. https://doi.org/10.1016/S0079-7421(08)60463-6

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science*, *21*(3), 415–423. https://doi.org/10.1177/0956797610362646

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 18, pp. 49–94). Academic Press.

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 444–458. https://doi.org/10.1037/0278-7393.17.3.444

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418–439. https://doi.org/10.1037/0278-7393.7.6.418

Hu, M., & Nosofsky, R. M. (2022). Exemplar-model account of categorization and recognition when training instances never repeat. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(12), 1947–1969. https://doi.org/10.1037/xlm0001008

JASP Team. (2023). *JASP* (0.17.3) [Computer software]. https://jasp-stats.org/

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, *120*(3), 628–666. https://doi.org/10.1037/a0033180

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 316–332. https://doi.org/10.1037/0278-7393.32.3.316

Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 626–640. https://doi.org/10.1037/0278-7393.29.4.626

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22

Little, D. R., Wang, T., & Nosofsky, R. M. (2016). Sequence-sensitive exemplar and decision-bound accounts of speeded-classification performance in a modified Garner-tasks paradigm. *Cognitive Psychology*, *89*, 1–38. https://doi.org/10.1016/j.cogpsych.2016.07.001

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

Mareschal, D., Quinn, P. C., & Lea, S. E. G. (Eds.). (2010). *The making of human concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199549221.001.0001

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*(3), 679–690. https://doi.org/10.3758/s13428-010-0049-5

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 128–148. https://doi.org/10.1037/0096-1523.21.1.128

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*(1), 113–138. https://doi.org/10.1146/annurev.ps.35.020184.000553

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775–799. https://doi.org/10.1037/0278-7393.27.3.775

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 275–292. https://doi.org/10.1037/0278-7393.28.2.275

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (1st ed., pp. 40–64). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.003

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*(1), 43–71. https://doi.org/10.1037/0096-3445.118.1.43

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE*, *10*(7), Article e0120644. https://doi.org/10.1371/journal.pone.0120644

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104–114. https://doi.org/10.1037/0278-7393.10.1.104

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

Nosofsky, R. M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1988b). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*(1), 25–53. https://doi.org/10.1146/annurev.ps.43.020192.000325

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In A. J. Wills & E. M. Pothos (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211–233. https://doi.org/10.1037/0278-7393.18.2.211

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315. https://doi.org/10.1037/a0022494

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924–940. https://doi.org/10.1037/0278-7393.28.5.924

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt .1), 353–363. https://doi.org/10.1037/h0025953

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2, Pt. 1), 304–308. https://doi.org/10.1037/h0028558

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

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*(1), 1–14. https://doi.org/10.1002/bdm.681

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton Century Crofts.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. https://doi.org/10.1037/0033-295X.107.2.358

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87. https://doi.org/10.1016/0022-2496(64)90017-3

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*(3), 278–304. https://doi.org/10.1037/0096-3445.121.3.278

Smith, J. D. (2014). Prototypes, exemplars, and the natural history of categorization. *Psychonomic Bulletin & Review*, *21*(2), 312–331. https://doi.org/10.3758/s13423-013-0506-0

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 800–811. https://doi.org/10.1037/0278-7393.28.4.800

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2019). Belief bias is response bias: Evidence from a two-step signal detection model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 320–332. https://doi.org/10.1037/xlm0000587

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 3–11. https://doi.org/10.1037/0278-7393.28.1.3

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*(1), 28–50. https://doi.org/10.1016/j.jmp.2003.11.004

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *2*(5–6), 440–445. https://doi.org/10.1016/S0022-5371(63)80045-6

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *1960 IRE WESCON convention record* (pp. 96–104). Institute of Radio Engineers.

Worthy, D. A., Maddox, W. T., & Markman, A. B. (2008). Ratio and difference comparisons of expected reward in decision-making tasks. *Memory & Cognition*, *36*(8), 1460–1469. https://doi.org/10.3758/MC.36.8.1460

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*(3), 387–402. https://doi.org/10.3758/BF03193783

Zaki, S. R., & Nosofsky, R. M. (2007). A high-distortion enhancement effect in the prototype-learning paradigm: Dramatic effects of category learning during test. *Memory & Cognition*, *35*(8), 2088–2096. https://doi.org/10.3758/BF03192940

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398. https://doi.org/10.3758/BF03193416

Received January 24, 2024
Revision received January 14, 2025
Accepted January 15, 2025 ■