Psychonomic Bulletin & Review (2024) 31:2118–2136
https://doi.org/10.3758/s13423-023-02448-2

BRIEF REPORT

Reconciling categorization and memory via environmental statistics

Arjun Devraj[1] · **Thomas L. Griffiths**[1,2] · **Qiong Zhang**[3]

Accepted: 16 December 2023 / Published online: 16 February 2024

Abstract

How people represent categories and how those representations change over time is a basic question about human cognition. Previous research has demonstrated that people categorize objects by comparing them to category prototypes in early stages of learning but consider the individual exemplars within each category in later stages. However, these results do not seem consistent with findings in the memory literature showing that it becomes increasingly easier to access representations of general knowledge than representations of specific items over time. Why would one rely more on exemplar-based representations in later stages of categorization when it is more difficult to access these exemplars in memory? To reconcile these incongruities, our study proposed that previous findings on categorization are a result of human participants adapting to a specific experimental environment, in which the probability of encountering an object stays uniform over time. In a more realistic environment, however, one would be less likely to encounter the same object if a long time has passed. Confirming our hypothesis, we demonstrated that under environmental statistics identical to typical categorization experiments the advantage of exemplar-based categorization over prototype-based categorization increases over time, replicating previous research in categorization. In contrast, under realistic environmental statistics simulated by our experiments the advantage of exemplar-based categorization over prototype-based categorization decreases over time. A second set of experiments replicated our results, while additionally demonstrating that human categorization is sensitive to the category structure presented to the participants. These results provide converging evidence that human categorization adapts appropriately to environmental statistics.

Keywords Category learning · Prototype model · Exemplar model · Memory retrieval · Environmental statistics · Rational analysis

Computational models of categorization have played a key role in understanding the representations people use when learning categories. Two prominent models of categorization are the prototype model, which posits that a stimulus is categorized by comparing it to the prototype of each category (Posner and Keele,1968; Posner and Keele, 1970; Reed, 1972; Mervis and Rosch, 1981; Homa, Sterling, & Trepel, 1981), and the exemplar model, which asserts that a stimulus is categorized by comparing it with all of the objects for each category (Medin & Schaffer, 1978; Nosofsky, 1986). Earlier findings strongly favored the exemplar model (McKinley & Nosofsky, 1996; Medin & Schaffer, 1978; Shin & Nosofsky, 1992), but these studies focused on the final stage of learning and did not consider the progression of category learning over time. Smith and Minda (1998) studied the fits of the prototype and exemplar models of categorization to human performance on a word categorization task over the course of an experiment and found that although the exemplar model dominates in the end, the prototype model has a strong

Experiment 1 appeared in a paper in the non-archival conference proceedings of the Annual Meeting of the Cognitive Science Society in 2021. A pre-print of the manuscript is available at psyarxiv.com.
This work was supported by the CV Starr Fellowship awarded to Q.Z. by Princeton Neuroscience Institute, a start-up fund awarded to Q.Z. by Rutgers University–New Brunswick, and AFOSR FA9550-18-1-0077 awarded to T.L.G. Correspondence concerning this article should be addressed to Qiong Zhang (qiong.z@rutgers.edu).

✉ Qiong Zhang
qiong.z@rutgers.edu

[1] Computer Science Department, Princeton University, 35 Olden St, Princeton, NJ 08540, USA

[2] Psychology Department, Computer Science Department, Princeton University, 35 Olden St, Princeton, NJ 08540, USA

[3] Psychology Department, Computer Science Department, Center for Cognitive Science, Rutgers University – New Brunswick, 152 Frelinghuysen Rd, Piscataway, NJ 08854, USA

advantage in early stages of learning. This result is consistent with the proposal that individuals categorize using a simple rule-based approach in early stages of category learning and more exemplar-specific strategies in later stages (Nosofsky, Palmeri, & McKinley, 1994a).

While prototype- and exemplar-based categorization have been extensively studied in the literature, less attention has been paid to the constraints of memory on categorization. How much one is able to use prototype-based or exemplar-based categorization is constrained by how accessible these representations are in memory. Prototype-based categorization requires retrieving generalized knowledge or summary statistics about the category, whereas exemplar-based categorization requires retrieving memory about specific instances. Categorizing using generalized knowledge in early stages and switching to an exemplar-based model in later stages, as documented by Smith and Minda (1998), does not seem consistent with findings in the memory literature showing that it becomes increasingly easier to access representations of general knowledge than representations of specific items over time – i.e., memory of specific items decays much faster than memory of the gist, or general knowledge (Posner and Keele, 1970; Zeng, Tompary, Schapiro, & Thompson-Schill, 2021). Why would one rely more on exemplar-based representations in late stages of categorization when it is more difficult to access these exemplars in memory?

Memory of specific items decays quickly for a reason. From a rational perspective, there is no need to retain information in memory if it is no longer needed, as determined by the statistics of the environment (Anderson & Schooler, 1991). Therefore, memory decay obeys a power-law function, defined as

$$f(x) = ax^k, \tag{1}$$

where $k < 0$ and x represents the amount of time that has passed, because the relationship between stimulus recency and the need odds, the odds that the stimulus will be required in the future, often obeys the power law in realistic environments, as determined from environmental sources such as *The New York Times*, parental speech, and electronic mail (Anderson and Schooler, 1991; Fig. 1A). In other words, we remember what happened five minutes ago better than what happened five days ago because what happened five minutes ago is more likely to be relevant to the current moment than what happened five days ago. In contrast, categorization experiments, including Smith and Minda (1998), usually use stimuli that appear with the same frequency regardless of the last time the stimulus was seen (Fig. 1B). This creates unrealistic environmental statistics where what happened five minutes ago is equally likely to be relevant to the current moment compared with what happened five days ago. To adapt to this new environment, human participants would over-represent stimuli in their memory that were last seen a long time ago.

To systematically examine the effect of environmental statistics on categorization behavior, we create an experimental condition in Experiment 1 that better reflects the power-law relationship between stimulus recency and the need odds as seen from environmental statistics encountered in the real world (Fig. 1C, D). We also create a control condition in which stimuli are presented with uniform frequency over time regardless of the last time a stimulus was seen, identical to Fig. 1B, following closely the setup in Smith and Minda (1998). We hypothesize that setting up the categorization experiment with realistic environmental statistics in the experimental condition will reduce the accessibility of exemplar-based representations in memory over time (assuming that human memory rationally adapts to the environment), therefore reducing or inverting the trend previously observed in favoring prototype-based representations early on and exemplar-based representations at late stages.

The goal of the current work is to demonstrate how human categorization behavior is a result of rationally adapting to the structure of the environment – different trajectories of prototype-based versus exemplar-based representations will emerge under different environmental statistics. Our efforts in providing a rational account of categorization regarding the environmental structure are in the same spirit as previous work that provides a rational account of categorization regarding the category structure, showing that when a learner should choose to use prototype-based or exemplar-based representations is determined by the category structure presented to the learner (Griffiths, Canini, Sanborn, & Navarro, 2007; Briscoe and Feldman, 2011). Prototype-based or exemplar-based representations correspond to different strategies of categorization, both being specific cases of a unifying model that is capable of optimally representing a given category structure (Griffiths et al., 2007). Similarly, it has been suggested that the literature that favors exemplar-based representations typically involves experiments that use poorly differentiated category structures (Medin, Dewey, & Murphy, 1983; Medin and Schaffer, 1978; Medin and Schwanenflugel, 1981; Medin and Smith, 1981; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994b; Nosofsky, Palmeri, and McKinley, 1994c); those category structures can weaken the urge to form prototype-based clusters of exemplars (Smith & Minda, 1998). Given the important role of category structure during categorization, in Experiment 2, we explore whether the results obtained in Experiment 1 can be generalized to different category structures by changing the number of category exceptions. We expect that when there are more category exceptions

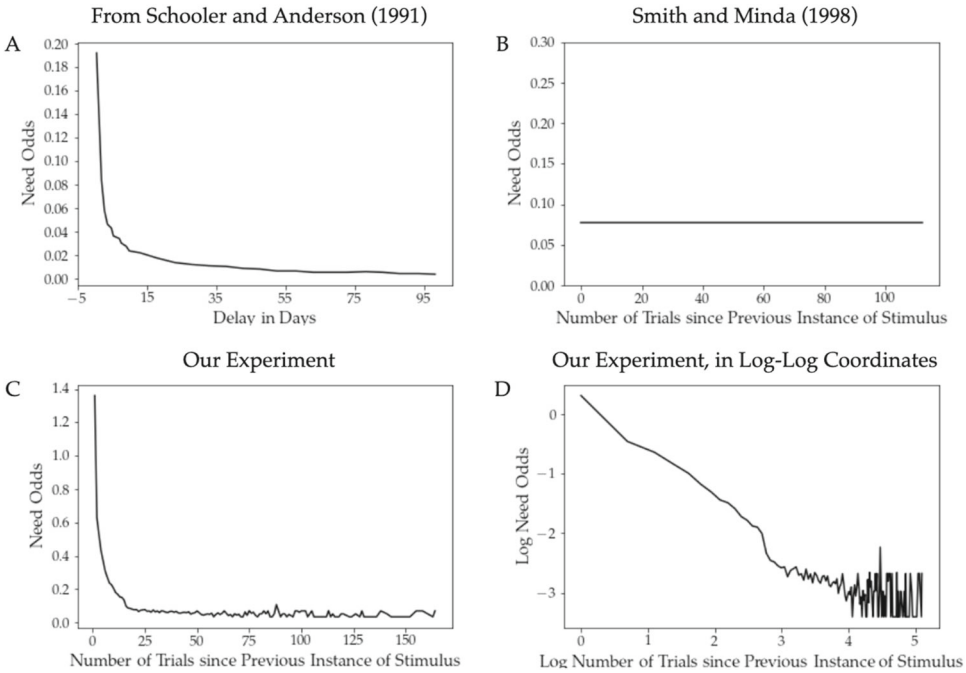

Fig. 1 The environmental recency function, displaying the relationship between stimulus recency and its need odds, calculated as $P(\text{stim})/(1-P(\text{stim}))$ where $P(\text{stim})$ is the probability of seeing the stimulus. **A** Based on words appearing in *The New York Times* headlines (1986 and 1987), reproduced from Fig. 4a in Schooler & Anderson (1991). **B** Calculated based on the experimental design of Smith and Minda (1998) and computed once every 14 trials using a 14-trial window. **C**, **D** Average data from the 60 sequences of stimuli generated for the experimental condition in Experiment 1. In **C** and **D**, $P(\text{stim})$ was calculated using a window size of 15 trials. The linear log-log relationship in **D** indicates that **C** is a power-law function, as taking the log of both sides of Eq. 1 gives the linear function $\log(f(x)) = k\log(ax)$

(non-linearly separable category structure), it is rational for participants to rely more on exemplar-based representations to encode these exceptions; whereas when there are no category exceptions (linearly separable category structure), prototype-based representations and exemplar-based representations are equally useful in supporting categorization performance. Of particular interest is how participants' categorization strategies over different stages of learning, as examined in Experiment 1, interact with different category structures.

The plan of the paper is as follows. We first explain the formulations of the prototype and exemplar models of categorization. Next, we delineate how we constructed the presentation of stimuli in the experimental condition in order to reflect the environmental statistics of the real world. Finally, we discuss the behavioral and model results from both our replication of Smith and Minda (1998) in the control condition and our manipulations of the environmental statistics in the experimental condition. Lastly, we conduct a second set of experiments to examine how our findings regarding the effect of the environmental statistics generalize to different category structures. These results provide converging evidence that human categorization adapts appropriately to environmental statistics.

Background

Prototype model

Various versions of the prototype model exist (Medin & Schaffer, 1978; Reed, 1972); we use here the formulation specified by Smith and Minda (1998, 2011). The prototype model compares the observed stimulus to the prototype for each category. First, the distance d_{i,P_k} between stimulus i and the prototype for category $k \in \{1, 2\}$, P_k, is computed as

$$d_{i,P_k} = \sum_{j=1}^{m} w_j |i_j - P_{k,j}|, \tag{2}$$

where i_j is the stimulus' value for dimension j, and $P_{k,j}$ is the value of the prototype of category k for dimension j, and w_j is the attentional weight assigned to dimension j. Each attentional weight w_j is constrained to take on a value in [0, 1], and the weights together sum to 1 ($\sum_{j=1}^{m} w_j = 1$).

Once the raw distance d_{i,P_k} has been calculated, stimulus i's similarity, η_{i,P_k}, to the category k prototype P_k is computed as

$$\eta_{i,P_k} = e^{-c\, d_{i,P_k}}, \tag{3}$$

where c is a sensitivity parameter constrained to [0, 20]. The sensitivity parameter has the effect of amplifying or shrinking psychological space (Smith & Minda, 1998). Finally, the probability that stimulus i (S_i) will be categorized into category 1 (R_1) is given by

$$P(R_1|S_i) = \frac{\eta_{i,P_1}}{\eta_{i,P_1} + \eta_{i,P_2}}. \quad (4)$$

Exemplar model

The exemplar model has been developed and generalized from the original context model (Medin, 1975; Nosofsky, 1984, 1986, 1987, 1988a; Palmeri & Nosofsky, 1995a; McKinley & Nosofsky, 1995a). This formulation of the exemplar model was used and thoroughly described in Smith and Minda (1998). The exemplar model compares the observed stimulus to all of the previously seen exemplars in each category in order to generate a category prediction for the observed stimulus. The distance $d_{i,x}$ and similarity measure $\eta_{i,x}$ between stimulus i and exemplar x are calculated identically as in the prototype model. However, the exemplar model considers all exemplars in order to generate a prediction, so the probability that stimulus i (S_i) will be categorized into category 1 (R_1) is

$$P(R_1|S_i) = \frac{\sum_{x\in C_1} \eta_{i,x}}{\sum_{x\in C_1} \eta_{i,x} + \sum_{x\in C_2} \eta_{i,x}}, \quad (5)$$

where C_1 is the set of exemplars for category 1 and C_2 is the set of exemplars for category 2.

Because our experiment used stimuli with six dimensions, both models had a total of six free parameters: the six attentional weights (w_1, ..., w_6) which together sum to 1 (hence, five of which are free) and the sensitivity parameter (c). The two models were then fit to participant data from the control and experimental conditions of the experiment. In addition to the versions of prototype and exemplar models described above, the major conclusions in the present work are also verified against other variants of models with a guessing-rate parameter (Smith & Minda, 1998), a response-scaling parameter (Nosofsky & Zaki, 2002), and exemplar memory-strength parameters (Donkin & Nosofsky, 2012). More details on these model variants can be found in Appendix A.

Experiment 1

Experiment 1 explores the possibility that participants' categorization strategies over different stages of learning depend on the environmental statistics. To this end, participants' performance is modeled using the versions of the prototype and exemplar models discussed earlier. In the control condition, set up similarly to Experiment 2 of Smith and Minda (1998), we expect to replicate the previous results in which the prototype-based model would fit human data better early on and the exemplar-based model would fit human data better later in learning; in the experimental condition with realistic environmental statistics, we predict that the trend observed in the control condition would be reduced or inverted.

Methods

Procedure

The task required participants to repeatedly categorize 14 stimuli into two categories. The stimuli were six-letter nonsensical words taken from Appendix A of Smith and Minda (1998). Each stimulus can be thought of as a six digit string of bits, such that the prototype for category 1 was `000000` and the prototype for category 2 was `111111`. Seven stimuli belonged to category 1 [`000000`, `100000`, `010000`, `001000`, `000010`, `000001`, `111101`], and the other seven stimuli belonged to category 2 [`111111`, `011111`, `101111`, `110111`, `111011`, `111110`, `000100`]. Each digit and position in the binary string corresponds to a unique letter; for example, the actual stimulus corresponding to `000000` that participants see is `gafuzi`, and the stimulus corresponding to `010000` is `gyfuzi`. In each trial, the participant was shown one of the 14 stimuli and had unlimited time to select a response of *1* for category 1 and *2* for category 2. For exactly 1 s after each trial, the participant was shown *Correct* or *Incorrect* accordingly.

Control condition The control condition, similar to Smith and Minda's (1998) experiment (the only difference being 616 trials in our experiment vs. 560 trials in the original experiment), consisted of 616 trials divided into 44 blocks, each consisting of 14 trials. Each block consisted of a random permutation of the 14 stimuli such that each stimulus was shown exactly once per block.

Experimental condition To create the environmental statistics in Fig. 1A, we developed an algorithm to generate a sequence of stimuli for the experiment. The specific pattern of stimulus presentation should, in turn, influence the need odds and hence the participant's retention of objects in memory. The experimental design first involved assigning each of the 14 stimuli its own power-law function. We added several new parameters to the power-law function to generate the necessary experimental conditions:

$$f(t) = \begin{cases} 0 & t < t_0 \\ s & t = t_0 \\ a(\frac{t-t_0}{r})^k - \theta & t_0 < t \le t_0 + n \\ 0 & t > t_0 + n \end{cases} \quad (6)$$

where t is the trial number, t_0 is the start trial for the stimulus to be introduced (i.e., an integer multiple of 35, since a new stimulus is introduced every 35 trials), r is the range, n is the number of trials the object's power-law function should be sampled from, s is the starting value for the power-law function, and θ is a calibration parameter. Power-law functions for consecutive stimuli were introduced in intervals of 35 trials. As in the control condition, each participant participated in a total of 616 trials. Figure 2A displays the resulting power-law functions for each stimulus over the entirety of the experiment. The assignment of stimuli to power-law functions was randomized for each participant: there was no balancing of exemplars from the two categories or any special treatment for the category exceptions. A different randomization was provided for each participant in order to ensure that the ordering of stimuli, such as whether the exceptions occurred early versus late, would not confound our results.

To assign a stimulus to each trial based on the continuous power-law functions, we first binned trials into discrete bins of 35 trials. (Since 616 trials is not divisible by 35, the final bin only contained data from 21 trials, though by this point all stimuli have already been introduced for a while.) We then generated a probability distribution for the bin based on the relative values of the various stimulus power-law functions over the trials in the bin, and generated the stimuli for trials in the bin by sampling from this distribution. Figure 2B shows how the resulting sequences of stimuli adequately reflected the original power-law functions. Figure 2C shows an example stimulus sequence for a single participant. Finally, we plotted how the need odds vary as a function of stimulus recency given the sequences of stimuli generated by this method (Fig. 1C, D), and verified that they closely reflect the environmental statistics in the real world such as those derived from headlines of *The New York Times* (Fig. 1A). Anderson and Schooler (1991) analyzed the statistics of real world environments such as *The New York Times* headlines, parental speech, and electronic mail, and found that the probability of encountering the same word again decreases over time. Following these patterns, the presentation structure in the experimental condition has the property that the spacing between presentations of an item tends to expand. The presentation structure has an additional feature – there are fewer items to sample from early on than later, similarly to how in a list-learning paradigm there are fewer items to rehearse from early on than later. While the first feature is necessary, it is possible that there exist other presentation structures without the second feature that can produce the environmental statistics analyzed by Anderson and Schooler (1991). Our goal here is to create an experimental setup (without exhausting all possible setups) that closely reflects the environmental statistics in the real world, as derived from *The New York Times* headlines, parental speech, and electronic mail.

Participants

We collected data from 60 participants for each condition (Control, Experimental). For reference, Smith and Minda (1998) had 32 participants for each experiment. Participants were recruited from the 18–23 age range and English-speaking population using Prolific.

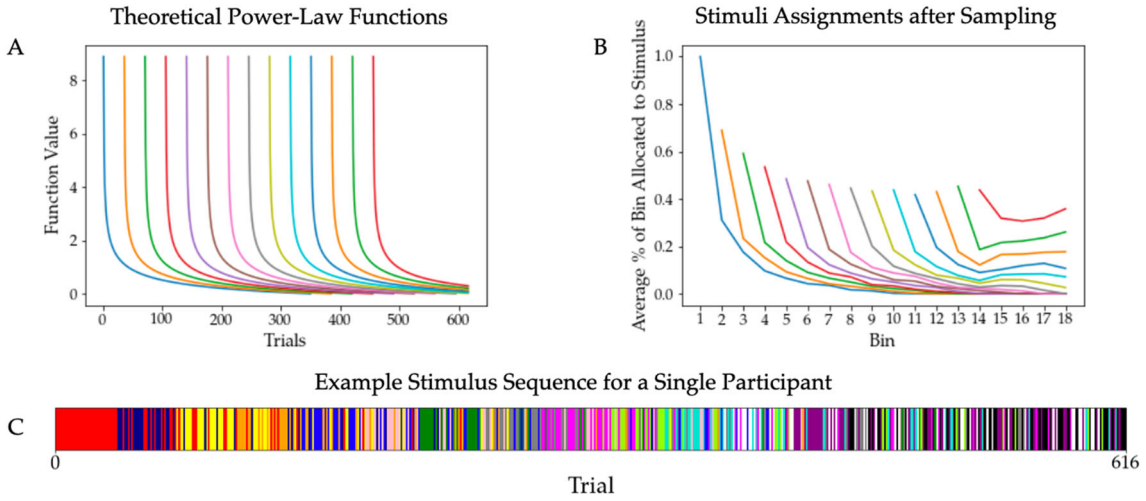

Fig. 2 The stimulus sequences shown to participants were generated algorithmically from power-law functions. **A** Each *curve* corresponds to the power-law function for a particular stimulus. **B** Since data is aggregated, each *curve* corresponds to the i-th stimulus seen across all participants. For example, the first curve illustrates the average proportion of each bin allocated to the first stimulus (regardless of its identity) shown to the 60 participants. **C** Each *color* corresponds to a unique stimulus $i \in \{1, 14\}$, and the figure highlights the significant temporal clustering and subsequent decay caused by our stimulus generation algorithm based on power-law functions. Parameter values of $a = 1$, $k = -0.3$, $r = 1000$, $n = 350$, $s = 10$, and $\theta = a * (\frac{n}{r})^k$ were chosen to adequately match the power-law function in Fig. 1A

Model fitting

Trials were split and grouped into 11 trial segments. For the control condition, following the procedure in Smith and Minda (1998), the first trial segment corresponds to trials at the beginning of the experiment, and the last trial segment corresponds to trials towards the end of the experiment. For the experimental condition, the first trial segment corresponds to the earliest stage when a stimulus has just been introduced (i.e., at the beginning of each power-law function in Fig. 2A), and the last trial segment corresponds to the latest stage when a stimulus has been introduced for a long time (i.e., towards the end of each power-law function in Fig. 2A). Note that we can also interpret the early and late trial segments in the control condition as coming from the earliest or the latest stage after a stimulus has been introduced. There are 56 trials exactly for each trial segment in the control condition, and 56 trials on average for each trial segment in the experimental condition due to the probabilistic sampling process.

Figure 3 provides a visualization of the trial structure: in earlier trial segments (Trial Segment 1) at the beginning of the power-law functions, each stimulus is presented very frequently with a very short duration between consecutive presentations, whereas in later trial segments (Trial Segment 11) towards the end of the power-law functions, each stimulus is presented less frequently with a larger gap between consecutive presentations.

To compare the control condition of our results with Smith and Minda's (1998), we followed the same model fitting methods as Smith and Minda (1998) to minimize the sum of squared errors (SSE) between the observed and predicted probabilities:

$$SSE = \sum_{i=1}^{14} (P(R_1|S_i) - \hat{p}_{1,S_i})^2 \tag{7}$$

where $P(R_1|S_i)$ is the model-generated probability that stimulus i belongs to category 1 based on an entire trial segment (56 trials) of data, and $\hat{p}_{1,S_i}$ is the proportion of trials in the trial segment (out of those in which stimulus i was seen) in which the participant actually categorized stimulus i in category 1. We ran Scipy's Sequential Least Squares Programming (SLSQP) method with ten initial random configurations to obtain the best-fit parameters. The model was fit for each participant receiving the control condition over each trial segment, as in Smith and Minda (1998). The resulting best-fit parameters were used to calculate the SSE that was ultimately used as the measure of fit in Fig. 5.

For the experimental condition, trial segments (as shown in Fig. 3) are not chronological groupings of trials in absolute time. This means that it is possible that two trials that are adjacent in time in the experimental condition come from two very different trial segments like 2 and 9, and it seems unrealistic to assume that one's attentional weights would alter so quickly from one trial to the next. Therefore, the model in the experimental condition was fit for each participant over the entire experiment, instead of over each trial segment like that done in the control condition. Additionally, because the number of trials per trial segment is not fixed in the experimental condition (due to the probabilistic sampling procedure for constructing the trial structure as shown in Fig. 2), we used the mean-squared error (MSE) over all trials in the trial segment, with residuals computed on a trial-by-trial basis rather than summed over an entire trial segment like that done in the control condition:

$$MSE = \frac{1}{56} \sum_{j=1}^{56} (P(R_1|T[j]) - \hat{p}_{1,T[j]})^2 \tag{8}$$

where j is the trial number, $T[j]$ corresponds to the stimulus S_i that was seen on trial j of the trial segment, and $\hat{p}_{1,T[j]}$ is

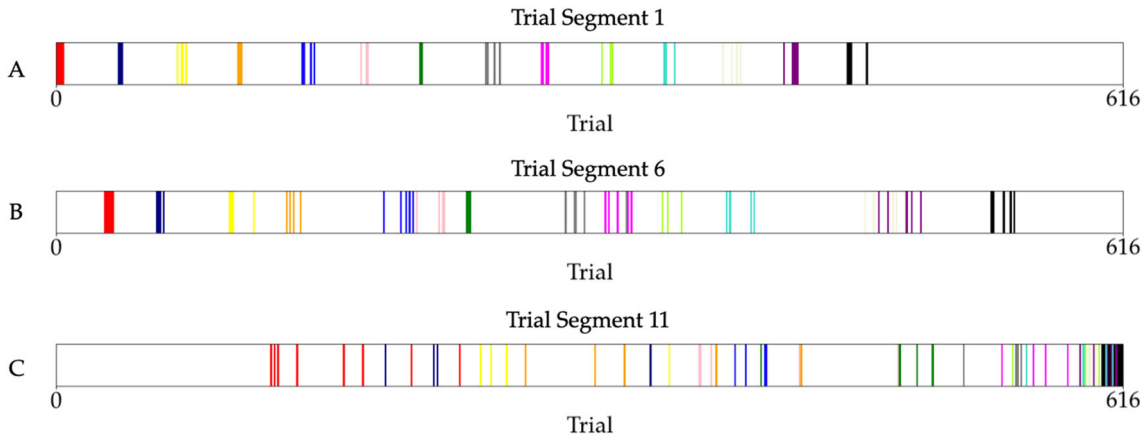

Fig. 3 An example of the sets of trials that constitute trial segments **A** 1, **B** 6, and **C** 11 in the experimental condition

either 0 or 1 depending on the participant's observed prediction on trial j.

We ensured that the exemplar model only considered exemplars seen *thus far* since not all exemplars were seen from the beginning in the experimental condition. That is, referencing Eq. 5 (which delineates the exemplar model), the summation of similarities between the stimulus and every exemplar in each category only incorporates those exemplars that the participant has already encountered (since during the experimental condition, some exemplars may only be seen for the first time very late in the experiment). Specifically, this means that for the exemplar model on a given trial, the C_1 and C_2 terms are not necessarily the entire set of exemplars for that category but instead, the set of exemplars already encountered for that category. To make sure that differences between model results in the control condition and the experimental condition do not entirely come from different model-fitting choices, we carried out additional model analyses when both models are fit to the entire experiment (Appendix A) and when both models are fit to consecutive sequences of 56 trials, as in the original control condition (Appendix B), which did not change our conclusions.

Results

Behavioral results

The categorization accuracies per trial segment in each condition were averaged over all participants and graphed over time, resulting in Fig. 4. In the control condition, significant learning occurred between the first ($M = .62, SD = .11$) and last ($M = .73, SD = .18$) trial segment as per the Wilcoxon signed-rank test, $z = 276.5$, $p < .001$. In the experimental condition, under realistic environment statistics where a stimulus is presented less and less frequently over trial segments, categorization accuracy decreases from the second ($M = .92, SD = .08$) to last ($M = .81, SD = .15$) trial segment as per the Wilcoxon signed-rank test, $z = 47.0$, $p < .001$. The second trial segment was compared in the experimental condition because this was the point that best reflected the beginning of the power-law function – when the participant had repeatedly seen the stimulus and was past the very initial stage of learning. In the control condition, where the probability of encountering a stimulus stays uniform over time, categorization accuracy increases as there is an opportunity to overlearn the material. In the experimental condition with realistic environmental statistics, accuracy decreases because from a rational perspective, it is not necessary to maintain strong memory representations of exemplars when the chance of needing them is low. One might wonder if this decrease in accuracy in the experimental condition is instead attributed to the number of stimuli introduced (since later trial segments tend to contain a greater diversity of stimuli on average than earlier ones). We ran an additional analysis that only considered trials in the experimental condition after all stimuli had been introduced and found that the decrease in accuracy between the second and last trial segment in the experimental condition is still statistically significant ($z = 380$, $p < .001$).

Model results

The model results over trial segments of the control condition in Fig. 5B are compared with the results from Smith and Minda (1998) in Fig. 5A. We qualitatively replicate the key results from Smith and Minda (1998), where the advantage of the exemplar model over the prototype model is increasing over time. The model results over trial segments of the experimental condition (Fig. 5C) are compared with the results from the control condition, when also fit over the entire experiment and using MSE for fit. In contrast to the control condition, under environmental statistics that reflect real-world environments, the advantage of the exemplar model over the prototype model is decreasing over trial segments in the experimental condition. To test the interaction of model type and trial segment as within-subject variables on model

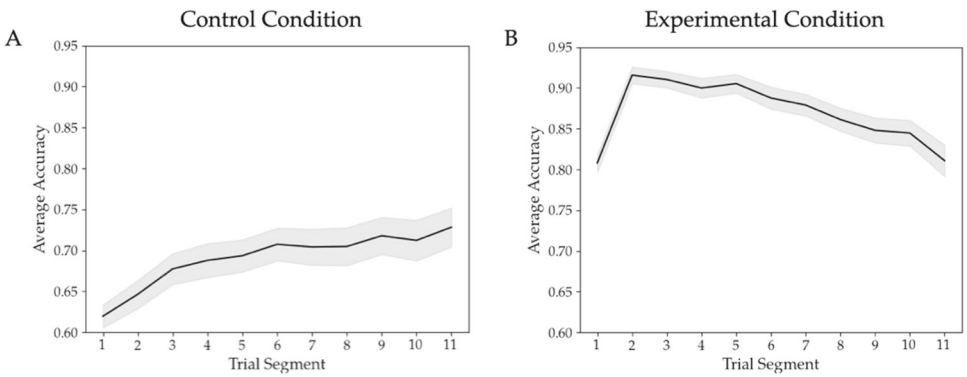

Fig. 4 Average categorization accuracy over trial segments for A the control condition and B the experimental condition. *Error bars* denote the standard error of the mean

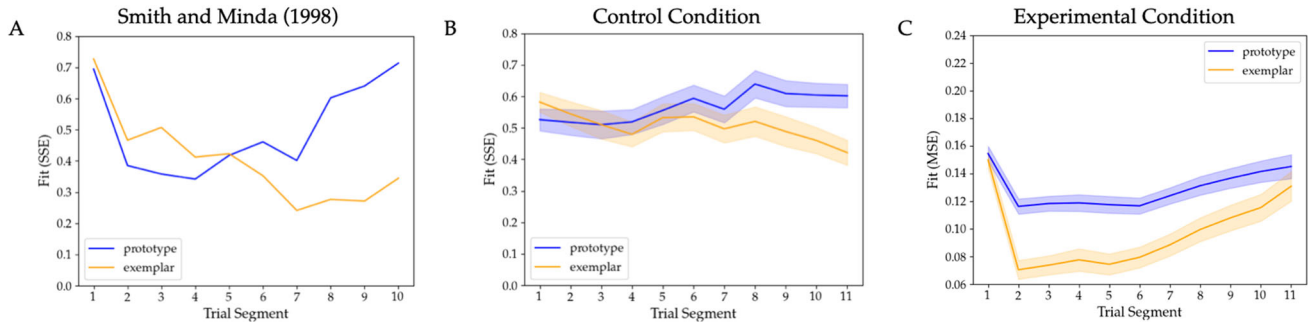

Fig. 5 Comparison of model fits of prototype-based versus exemplar-based categorization for **A** the original experiment, reproduced from Smith and Minda (1998), **B** the control condition, and **C** the experimental condition after introducing realistic environmental statistics. Models were fit for each participant's performance over a trial segment (**A**, **B**) or over the entire experiment (**C**) and then averaged over all participants for the condition. *Error bars* denote the standard error of the mean

fit, we randomly permuted the model type and trial segment variables within each subject and computed the resulting interaction F test statistic. In this test, the p value is the percentage of simulations for which the interaction F test statistic exceeds the interaction F test statistic from the real data. By running 1000 simulations, we report $p < .001$ in the test for interaction between model type (prototype, exemplar) and trial segment (first/second, last) in both the control and experimental conditions of Experiment 1. We conducted additional model simulations in Appendix E, using either the prototype or exemplar model as the ground truth, and plotted corresponding model fits similar to Fig. 5. While the relative fits of our model consistently reflect the ground-truth model, we observed that the change in fits of the exemplar-based model (but not the prototype-based model) is influenced by the number of stimuli encountered thus far. This effect could contribute to the trend depicted in Fig. 5. Therefore, we further analyzed periods of the experiment when the number of stimuli that have been seen is fixed, and demonstrated that the trend persists (see details in Appendix E). More details of the distribution of responses over stimuli with what the prototype model and the exemplar model predict can be found in Appendix D.

Experiment 2

In addition to environmental statistics, Experiment 2 also assessed the effect of category structure on the fits of the prototype and exemplar models by altering the number of category exceptions (i.e., 0, 2, 4). Since the prototype model is equivalent to assuming a linear boundary between categories (Ashby & Maddox, 1993), we expect that when there are more category exceptions it is rational for participants to rely more on exemplar-based representations to encode these exceptions; whereas when there are no category exceptions participants would rely similarly on both prototype-based or exemplar-based representations. Of particular interest is how participants' category strategies over different stages of learning depends jointly on the category structure and environmental statistics. Since Experiment 1 was conducted using a category structure with two exceptions, we additionally expect to replicate the results in Experiment 1 in the conditions with two exceptions in Experiment 2.

Methods

Procedure

Experiment 2 has a total of six conditions. We examine category structures with 0 exceptions, two exceptions (as in Experiment 1), and four exceptions, with a control and experimental condition for each level of exceptions. The sequences of stimuli for each of the experimental and control conditions were generated identically as in Experiment 1, using the same binning and sampling method with the piece-wise power law function in Eq. 6 for the experimental condition and repeated blocks of random permutations of the 14 stimuli for the control condition; in comparison to Experiment 1, only the actual stimuli changed, depending on the number of exceptions. Stimuli were still six-letter nonsensical words modeled by six-bit binary strings, with the category prototypes of `000000` for category 1 and `111111` for category 2. For the conditions with 0 exceptions, the seven stimuli in category 1 were `000000`, `100000`, `010000`, `001000`, `000010`, `000001`, and `000100`, and the seven stimuli in category 2 were `111111`, `011111`, `101111`, `110111`, `111011`, `111110`, and `111101`. For the conditions with two exceptions, the category structure was identical to that described in Experiment 1: `000000`, `100000`, `010000`, `001000`, `000010`, `000001`, and `111101` for category 1 and `111111`, `011111`, `101111`, `110111`, `111011`,

111110, 000100 for category 2. For the conditions with four exceptions, the seven stimuli in category 1 were 000000, 100000, 010000, 000010, 000001, 111101, and 101111, and the seven stimuli in category 2 were 111111, 011111, 110111, 111011, 111110, 000100, and 001000. Each digit and position in the binary string corresponds to a unique letter; for example, the actual stimulus corresponding to 000000 that participants see is gafuzi, and the stimulus corresponding to 010000 is gyfuzi.

Participants

This experiment was run on Prolific with 60 participants for each of the six conditions. Exactly as in Experiment 1, participants were recruited from the 18–23 age range and English-speaking population.

Model fitting

The model-fitting procedure for both control and experimental conditions was identical to that in Experiment 1.

Results

Behavioral results

We replicated key behavioral results from Experiment 1: as shown in Fig. 6, in control conditions, participants improve their categorization accuracy over trial segments, whereas in experimental conditions, participants' categorization accuracy decreases over trial segments. Additionally, we examined the effect of exceptions on categorization accuracy. As the number of exceptions increases, categorization accuracy declines for both the control and experimental conditions. To formally test these effects, we used the participants' categorization accuracies with trial segment (first/second, last) as a within-subject variable and number of category exceptions (0, 2, 4) as a between-subjects variable for both the control and experimental conditions and conducted a permutation test with 1000 simulations, for which we computed the F test statistic for each independent variable. For both the control and experimental conditions, there were significant effects on categorization accuracy by both the number of exceptions, $p < .001$, and the trial segment, $p < .001$.

Model results

There are two goals for the modeling analyses. The first goal is to examine the effect of category structure (i.e., number of exceptions) on the fits of the prototype and exemplar models. Figure 7 shows that as the number of exceptions increases, the exemplar model's advantage over the prototype model also increases. We analyzed the effects of model type (prototype, exemplar), a within-subject variable, and number of exceptions (0, 2, 4), a between-subjects variable, on the model fits in the middle (fifth) trial segment by running 1000 simulations with random permutations of the model type and number of exceptions variables and computing the resulting interaction F-test statistic. Using an identical definition of p value as for the permutation tests in Experiment 1, the interaction between model type and number of exceptions was significant for both the control conditions, $p = .022$, and experimental conditions, $p < .001$.

The second goal of the modeling analyses is to examine whether the results obtained in Experiment 1 can be generalized to different category structures (linearly-separable: 0 exceptions; non-linearly separable: two exceptions, four exceptions). Consistent with the observations in Experiment 1, for conditions with two or four exceptions, the exemplar model's advantage over the prototype model decreases over time in the experimental condition but increases over time in the control condition, as shown in Fig. 7. However, for conditions without any exceptions, the model fits for both the prototype and exemplar models

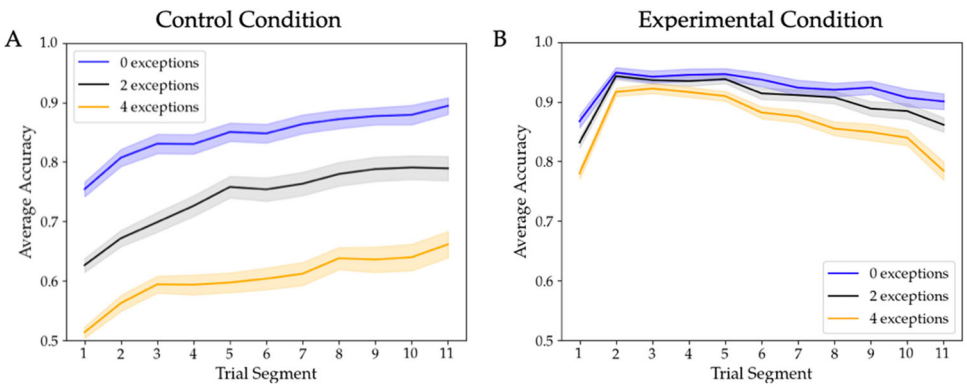

Fig. 6 Average categorization accuracy over trial segments for all three control conditions and all three experimental conditions, by number of exceptions. *Error bars* denote the standard error of the mean

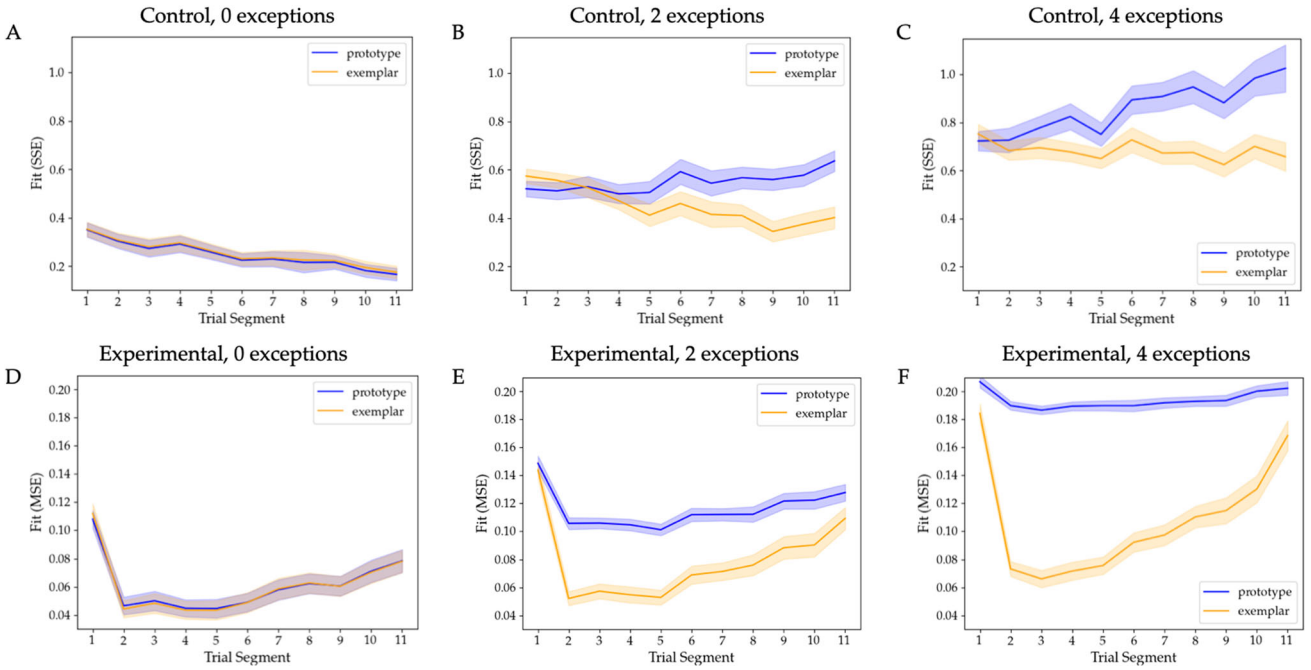

Fig. 7 Comparison of model fits of prototype-based versus exemplar-based categorization for all six conditions, each defined by (1) control or experimental version of the experiment and (2) the number of exceptions in the category structure. Models were fit for each participant's performance over a trial segment (**A**–**C**) or over the entire experiment (**D**–**F**) and then averaged over all participants for the condition. *Error bars* denote the standard error of the mean

exhibit nearly identical trends over time and do not demonstrate the effects previously observed in Experiment 1. This could be attributed to the similar predictions made by the exemplar-based model and the prototype-based model under this category structure. To formally test these effects, we ran a non-parametric permutation test using the F interaction test statistic to analyze the effects of model type (prototype, exemplar) and trial segment (first/second, last) on the model fits – identical to the test in Experiment 1. For the cases with zero exceptions, we obtain $p = .005$ for the experimental condition and $p = .048$ for the control condition; for the cases with two exceptions and four exceptions, we obtain $p < .001$ for both the control and experimental conditions.

We have shown that the overall accuracy is higher when there are fewer exceptions. This is because when there are no exceptions, prototype-based representations and exemplar-based representations are equally useful in supporting categorization performance. When there are category exceptions, accuracy critically depends on the ability to form exemplar-based representations, as exemplar-based representations are capable of encoding both exceptions and non-exceptions whereas prototype-based representations provide summary statistics representative of only non-exceptions. Similarly, the increasing accuracy over trial segments in the control condition in Fig. 6 can be explained by the improved fits of exemplar-based representations relative to prototype-based representations (control conditions in Fig. 7), when there is opportunity to over-learn the stimuli. The decreasing accuracy over trial segments in the experimental condition in Fig. 6 can be explained by the decreased fits of exemplar-based representations relative to prototype-based representations (experimental conditions in Fig. 7), when stimuli are presented less and less frequently.

General discussion

While prototype-based and exemplar-based categorization have been extensively studied in the literature, less attention has been paid to the constraints of memory on categorization under different environmental statistics. In our results, we demonstrate that human categorization rationally adapts to the environmental statistics. In our experiments, average categorization accuracy generally improved over time in the control condition; this pattern is expected, as participants were shown stimuli with equal frequency throughout the experiment and thus could learn from more information over time. In the experimental condition, however, categorization accuracy steadily declined after the second trial segment. Therefore, our experimental design was successful at introducing memory constraints that inhibited performance over time. In Experiment 1, our findings from the control condition

qualitatively replicated Smith and Minda's (1998) findings that when stimuli are presented with uniform frequency over time, there is an opportunity to over-learn the stimuli, which leads to an increased advantage of the exemplar model over the prototype model over time. The results from the experimental condition confirm our hypothesis: when stimulus presentation more accurately reflects realistic environmental statistics, there is reduced need to retrieve the same exemplars as time goes by, leading to a decreased advantage of the exemplar model over the prototype model. Experiment 2 replicated our results, while additionally demonstrating that human categorization is sensitive to the category structure presented to participants: the trend observed in adapting to different environmental statistics only extends to category structures with exceptions present. We now turn to the broader implications of these results.

It has been pointed out that memory and categorization rest on similar representations, but differ in whether identification is made at the instance/individual level or the category level (Anderson, 1991). By adjusting the level of identification, rational models of categorization can be used to explain results of fan effects in a memory task (Anderson, 1991). Exemplar-based models can account for connections between categorization and other fundamental cognitive processes such as in old-new recognition memory tasks (Estes, 1994; Hintzman, 1988; Nosofsky, 1988b, 1991; Nosofsky Zaki, 2003), as well as in short-term memory paradigms (Nosofsky, Little, Donkin, & Fific, 2011). Other studies in the literature have drawn parallels between category learning and memory studies. Following standard categorization experiments where participants made judgment at the category level, participants completed an additional recognition test that allows for probing the representations at the instance level. These studies identified a number of parallels between category learning and memory studies on the effects of schemas (Palmeri & Nosofsky, 1995a; Palmeri & Nosofsky, 1995b; Rojahn & Pettigrew, 1992; Sakamoto & Love, 2004).

The current paper is another step towards linking the two pillars of cognitive psychology: memory and categorization. Our findings demonstrate that what can account for rational memory behavior can also be used to explain categorization behavior. Since categorization requires not only retrieving category representations – which characterizes memory behavior – but also requires a step to use these representations to make an identification on the category, it is expected that what influences retrieval may also have downstream effects on the categorization behavior. Our results are consistent with the rational analysis of memory by Anderson (1990) in deciding whether one should retrieve a particular memory representation. A rationally designed information-retrieval system would retrieve memory structures ordered by their need probability p, and stop retrieving when $pG < C$, where G is the reward associated with retrieving a target memory and C is the cost in considering the memory. Under realistic environmental sources where p decreases as time goes by (Anderson and Schooler, 1991; Fig. 1A), this retrieval strategy predicts that information that has occurred more recently is more likely to be retrieved by the memory system (Anderson, 1990). Similarly, in our categorization experiments, under realistic environmental statistics where p decreases as time goes by (i.e., experimental condition), stimuli that have occurred further in the past are less likely to show up again and therefore less likely to be retrieved during categorization, leading to decreased access to exemplar-based representations over time. This contrasts with environmental statistics typically used in the categorization literature (i.e., control condition), where p stays constant as time goes by. The latter case provides an opportunity for over-learning individual stimuli, leading to increased access to exemplar-based representations over time regardless of when the stimuli were last seen. To summarize, our results from the control and experimental conditions in Experiments 1 and 2 are consistent with predictions from a rational-retrieval strategy when p is determined by the environmental statistics. Closely related to our hypothesis is a version of the exemplar model that directly imposes a power-law forgetting function; in this version, recent exemplars are more strongly weighted than remote ones during exemplar-based categorization (Elliott & Anderson, 1995; McKinley & Nosofsky, 1995b). Incorporating the forgetting function is critical to capture behavior in short and long-term recognition tasks and categorization experiments with changing definition of categories (Donkin & Nosofsky, 2012; Elliott & Anderson, 1995). Although the forgetting function also models the accessibility of exemplars over time, it does not embody the same assumption as in the present work. While an exemplar-based model with a forgetting function captures the relative access to one exemplar versus other exemplars, all exemplars can still be accessed, which is different than examining how much categorization could depend on a prototype-based representation instead as a result of having less access to the exemplars. On the other hand, we acknowledge that an exemplar-based model with a forgetting function is a more accurate exemplar-based model, especially in the experimental condition where the frequency of stimuli changes over time. Therefore, we additionally tested and verified our hypotheses under this variant of the exemplar-based model (see more details in Appendix A and B).

Furthermore, our results from Experiment 2 on the effect of category structure are consistent with predictions from a rational retrieval strategy when the reward G is determined by the category structure. Specifically, when there are no exceptions, prototype-based representations or exemplar-based representations are equally useful in supporting categorization performance, leading to the same G values. When there are category exceptions present, there is larger G associated with exemplar-based representations, as exemplar-based

representations are capable of encoding both exceptions and non-exceptions whereas prototype-based representations provide summary statistics representative of only non-exceptions. As a result, we observe that the advantage of exemplar-based representations over prototype-based representations increases as the number of exceptions increases, and that when there are no exceptions, there is no change in the difference of exemplar-based and prototype-based model fits over different stages of category learning.

While the memory literature shows that it becomes increasingly easier to access representations of abstract knowledge than representations of specific items over time (Posner & Keele, 1970; Zeng et al., 2021), categorization has been shown to rely on abstract knowledge in early stages and switch to an exemplar-based model in later stages (Smith & Minda, 1998). Our findings help to reconcile the incongruities between these results by demonstrating that different categorization behaviors arise as a result of rationally adapting to different environmental statistics. In both experiments, human participants rationally adjust to the need probability, relative cost, and reward associated with each object. Future work will explore more complex environmental statistics and cost-reward structures to better understand how memory and categorization work in tandem to enable efficient category learning in the real world.

Appendix

A. Models containing additional parameters, fitted for the entire experiment per participant

We explored additional versions of the prototype and exemplar models containing additional parameters as used in the literature. In particular, we introduced (1) the guessing-rate parameter, g, which measures the proportion of the time the participant was randomly guessing as opposed to actively using the model, (2) the response-scaling parameter, γ, which describes the degree to which the participant's choice is deterministic (Nosofsky & Zaki, 2002), and (3) each exemplar x's memory strength, $m_{x,j} = \alpha j^{-\beta}$, where α, the scaling parameter, can be assumed as 1, j is the lag (i.e., number of trials since exemplar x was last seen), and β indicates the rate of memory decay (Donkin & Nosofsky, 2012). While the guessing-rate parameter applies to both models, the response-scaling and exemplar memory-strength parameters only apply to exemplar model. After incorporating all new parameters, we arrive at the probability that stimulus i (S_i) will be categorized into category 1 (R_1) by the prototype model as

$$P(R_1|S_i) = \frac{g}{2} + (1-g) * \frac{\eta_{i,P_1}}{\eta_{i,P_1} + \eta_{i,P_2}} \tag{9}$$

and by the exemplar model as

$$P(R_1|S_i) = \frac{g}{2} + (1-g) * \frac{(\sum_{x \in C_1} m_{x,j} * \eta_{i,x})^{\gamma}}{(\sum_{x \in C_1} m_{x,j} * \eta_{i,x})^{\gamma} + (\sum_{x \in C_2} m_{x,j} * \eta_{i,x})^{\gamma}}. \tag{10}$$

We applied a similar approach for model-fitting and fit-calculation as with the experimental conditions of Experiments 1 and 2 and considered the effects of adding each new parameter individually as well as their combined effects. For model simplicity, we obtained the fixed value for the memory strength parameter $\beta = 1.4025$ from Donkin and Nosofsky (2012). We also tested that the specific choice of β value does not affect our main conclusions by additionally testing $\beta = 0.3$ and $\beta = 5$, as β values in previous studies fall in the range of 0.3 to 5 (Nosofsky, Cao, Cox, & Shiffrin, 2014a; Nosofsky, Cox, Cao, & Shiffrin, 2014b). We applied these new model versions to both the control and experimental conditions of Experiment 1 (i.e., two category exceptions). The fits of these new models, as shown in Fig. 8, follow the trends from the original models presented in this paper.

B. Models containing additional parameters, fitted for consecutive sets of 56 trials per participant

We also explored a model-fitting and fit-calculation approach that served as an intermediate between Smith and Minda's (1998) original approach (model-fitting per participant and per trial segment, plotted fit as the SSE over the trial segment) and our approach (model-fitting per participant over the entire experiment, plotted fit as the MSE over the trial segment after applying the model on a trial by trial basis). In the intermediate approach for *both* the control and experimental conditions, we fit the model for each participant's behavior for each consecutive set of 56 trials (i.e., how trial segments are defined in the control condition), hence generating 11 different fittings per participant, and then apply the model on a trial by trial basis for each trial segment and compute the MSE as the measure of fit. With this approach, one can still observe the trajectories of the two models over trial segments as defined in the experimental condition (i.e., temporally non-contiguous sets of trials that correspond to similar regions along each object's power-law function), but also take into consideration the potential for parameter values to change over the course of the experiment as Smith and Minda (1998) do. This means that in the experimental condition, for any given participant, the fitted parameters may differ even within a trial segment, since different trials may belong to different sets of 56 consecutive trials.

We applied the new models (as well as the intermediate models, derived by individually adding each of the new parameters) described in the previous section with this fit-

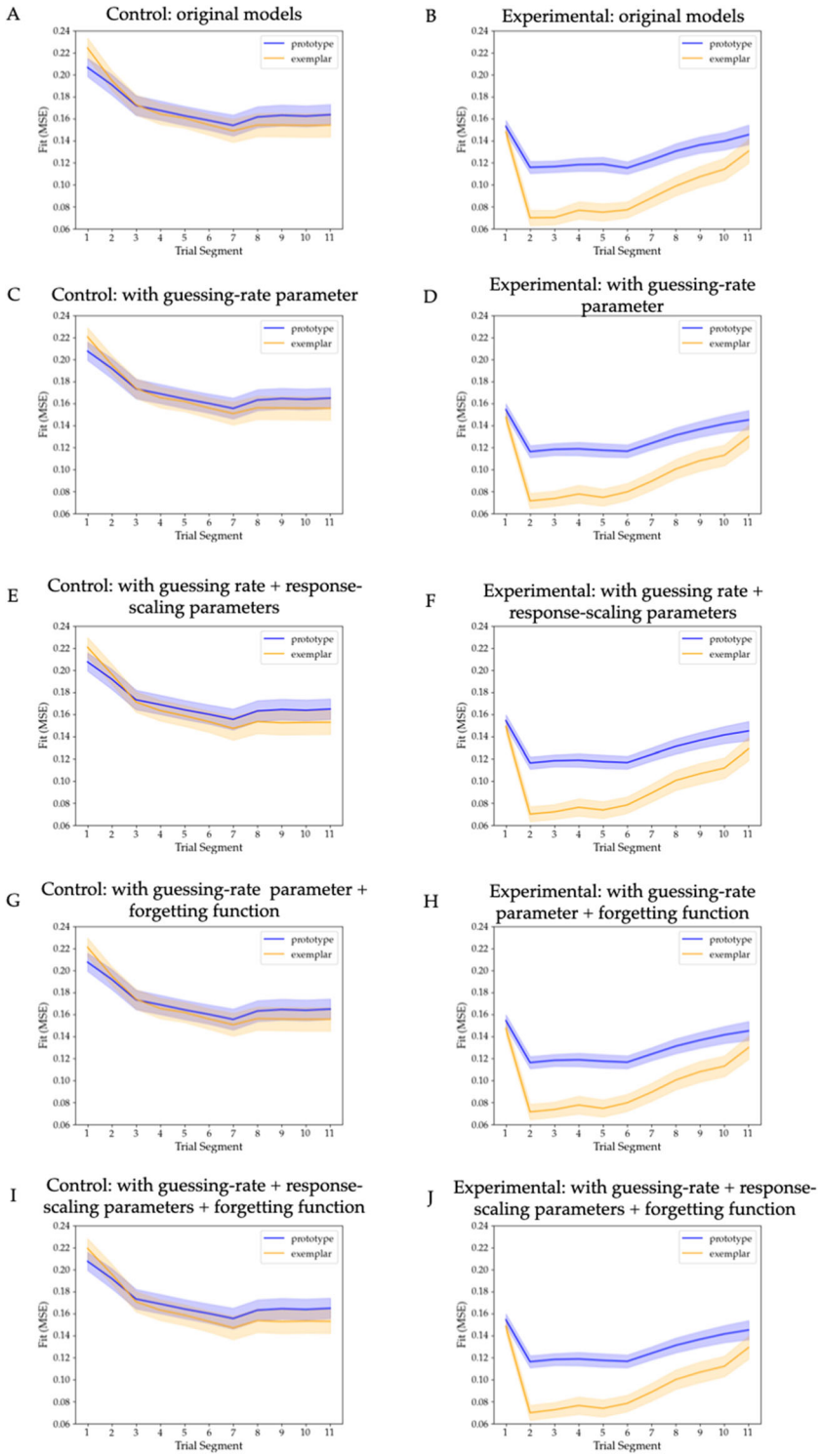

Fig. 8 Prototype and exemplar model fits with new parameters based on fitting the models per participant for the entire experiment

ting approach and obtained the results in Fig. 9. The patterns observed with our original versions of the models still hold, with the exemplar model gaining an advantage later in category learning in the control condition and losing its advantage in the experimental condition. In fact, after introducing the response-scaling parameter, we observe the first cases in which the prototype model actually outperforms the exemplar model in the final trial segments of the experimental condition, as seen in Fig. 9F and J. With this new model-fitting approach and using the same permutation test with

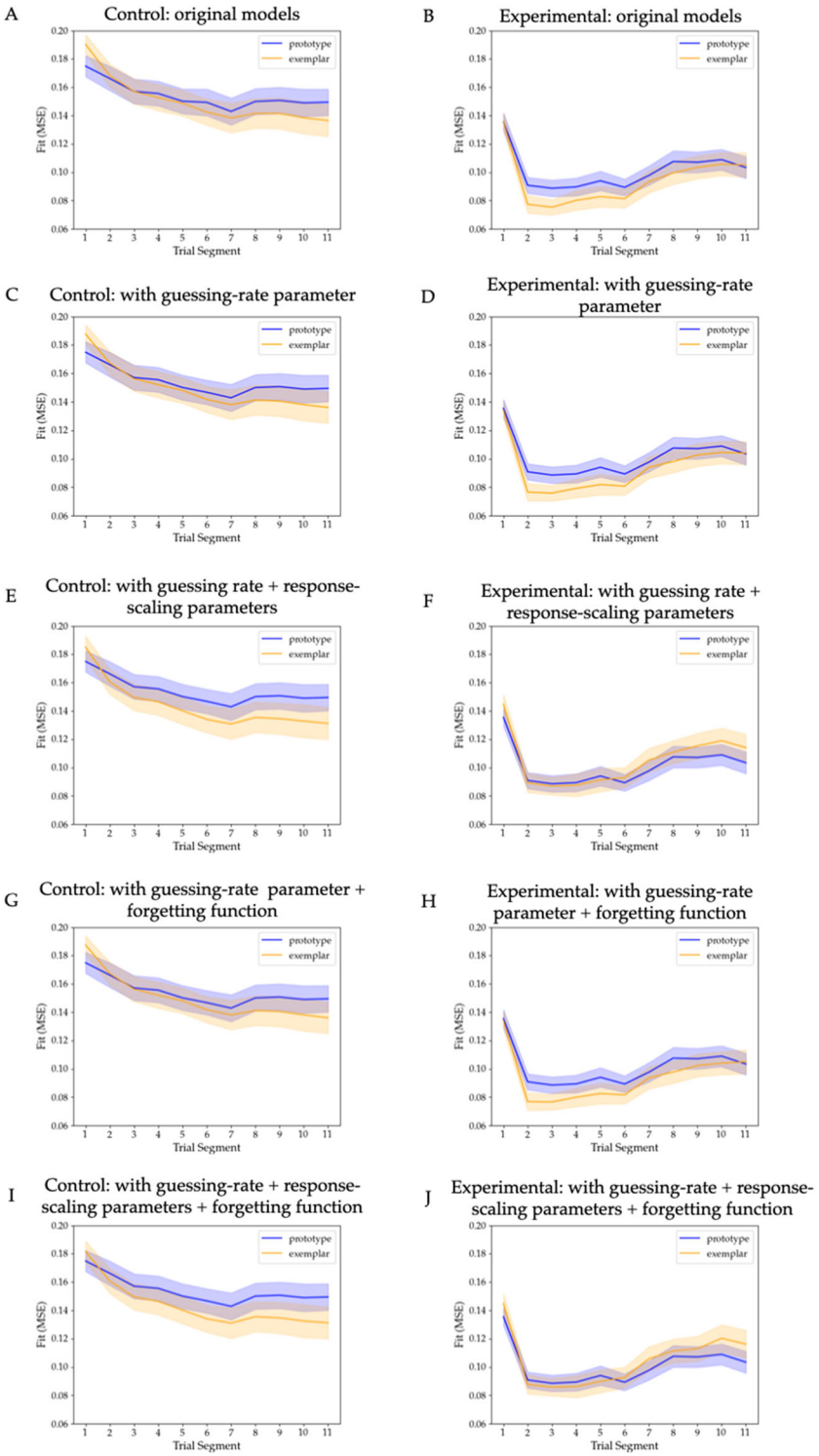

Fig. 9 Prototype and exemplar model fits with new parameters based on fitting the models per participant for consecutive sets of 56 trials

1000 random simulations as in Experiment 1, the interaction between model type and trial segment in the experimental condition for the model including all the new parameters (corresponding to Fig. 9J) was still statistically significant, $p = .007$.

C. Accuracy per stimulus in the experimental condition

We more closely analyzed the behavioral results for the experimental condition by breaking down participants' cate-

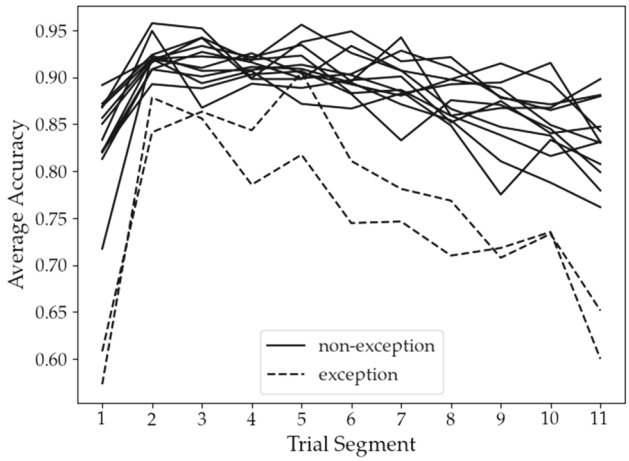

Fig. 10 Categorization accuracy per stimulus for the experimental condition of Experiment 1

gorization accuracy over trial segments by stimulus, as shown in Fig. 10. Since the trial segment in the experimental condition captures the relative frequency of stimulus presentation, Fig. 10 shows that (1) for every stimulus, regardless of its status as an exception/non-exception, the participants' categorization accuracy peaked early on (either in the second or third trial segment) and then declined afterwards, and (2) categorization accuracy for exceptions was generally lower throughout the experiment and also tended to decline more rapidly. These results are consistent with our hypothesis that people rationally adapt to the environmental statistics of the stimuli, and have worse representations of the stimuli when they are needed less (i.e., presented less frequently).

D. Comparison of model predictions to participant responses

While model fits and categorization accuracy are important metrics in our analyses, we also provide a more thorough account of the exact responses predicted by each model and how they compare to participants' responses in critical trial

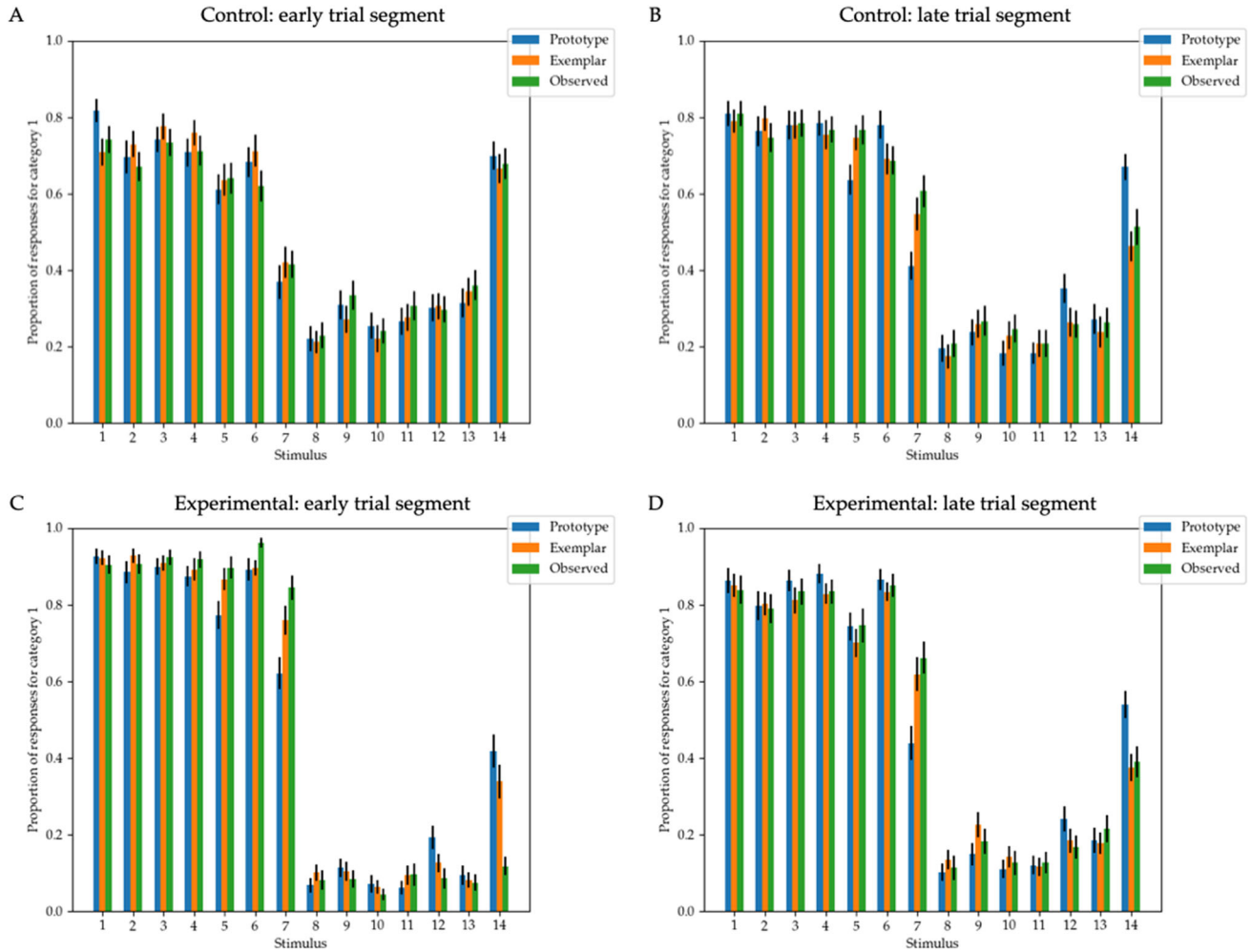

Fig. 11 Responses predicted by the prototype and exemplar models compared to participants' responses for critical trial segments

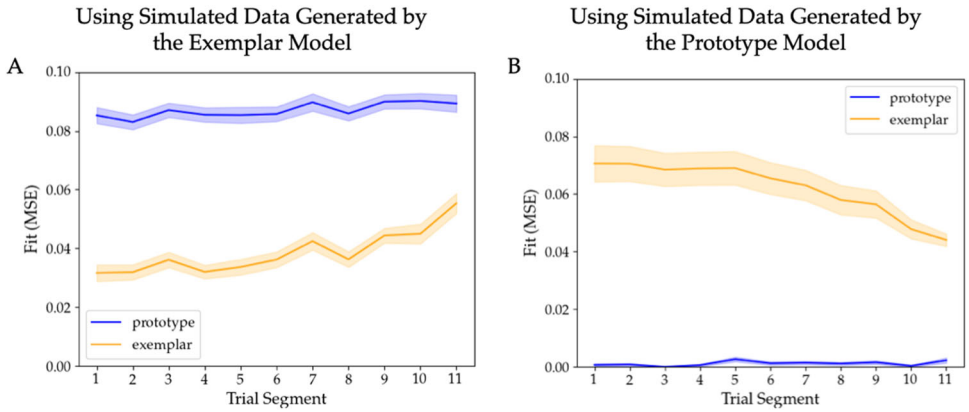

Fig. 12 Model fits for the experimental condition of Experiment 1 with model-generated simulated data

segments. For both conditions of Experiment 1, we analyzed responses for the second trial segment, when categorization accuracy tends to peak in the experimental condition, and the final trial segment, in order to capture the full range of behavior in the experiment. In particular, we measured the average proportion of trials that each of the models and participants selected category 1 for each distinct stimulus within each trial segment. We generated the model response by sampling from a Bernoulli distribution seeded with the fitted model's predicted probability for category 1. We used the model-fitting procedure in Appendix B, which fits the models per set of consecutive 56 trials. The results in Fig. 11 show that both the prototype and exemplar models captured the qualitative trends of the participants' responses reasonably well, across all 14 stimuli, in early/late trial segments of control/experimental conditions, though the prototype model struggled in predicting responses for exceptions (Stimulus 7 and Stimulus 14). To further understand quantitatively how well prototype models and exemplar models capture the participants' responses, we resort to the calculations of model fits, as illustrated in Fig. 5.

E. Confirming the results using simulated data

To better interpret the model fits, we produce simulations where the ground truth is known (i.e., either an exemplar model or a prototype model) and fit our models to the simulated data. Thus, for the exact stimulus sequences used in Experiment 1, we generated simulated response data by using the prototype or exemplar model as the ground truth (with each attentional weight set as $\frac{1}{6}$ and the remaining parameters fixed at the median of their allowed range, and then generating a Bernoulli sample for the trial from the model's predicted probability), fit the models to this data, and computed and graphed the model fits, as shown in Fig. 12.

Overall, our model-fitting procedure is robust to the ground-truth model, as we observed better fits (lower MSE) in the exemplar model than the prototype model when an exemplar model is the ground truth (Fig. 12A) and vice versa (Fig. 12B). However, the change of the model fits over trial segments only holds constant in the prototype model but not in the exemplar model, likely being affected by the total number of stimuli analyzed under a given trial segment

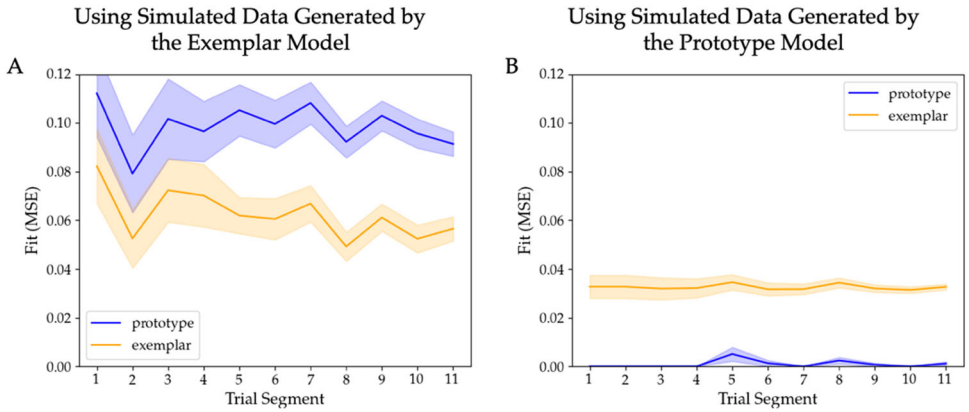

Fig. 13 Model fits for the experimental condition of Experiment 1 with model-generated simulated data, considering only trials after which all exemplars have already been seen

(later trial segments in the experimental condition contain a larger number of stimuli on average than earlier segments, as more of them have been encountered over time). To eliminate the effect introduced by the number of exemplars previously encountered, we consider only trials in the simulated data after which all stimuli have been encountered, which generally would occur around or slightly after trial 455, corresponding to $\sim$ 25% of the total experiment length. By analyzing this portion of the simulated data, model fits are no longer affected by the number of stimuli analyzed, as they all stay constant over trial segments under the ground truth of either an exemplar model or a prototype model (Fig. 13A, B).

Finally, we examine whether the main conclusion of our analysis of the empirical data in the experimental condition still holds after controlling for the number of exemplars. In this analysis, we only consider the sequence of trials between the introductions of every two consecutive stimuli so that the

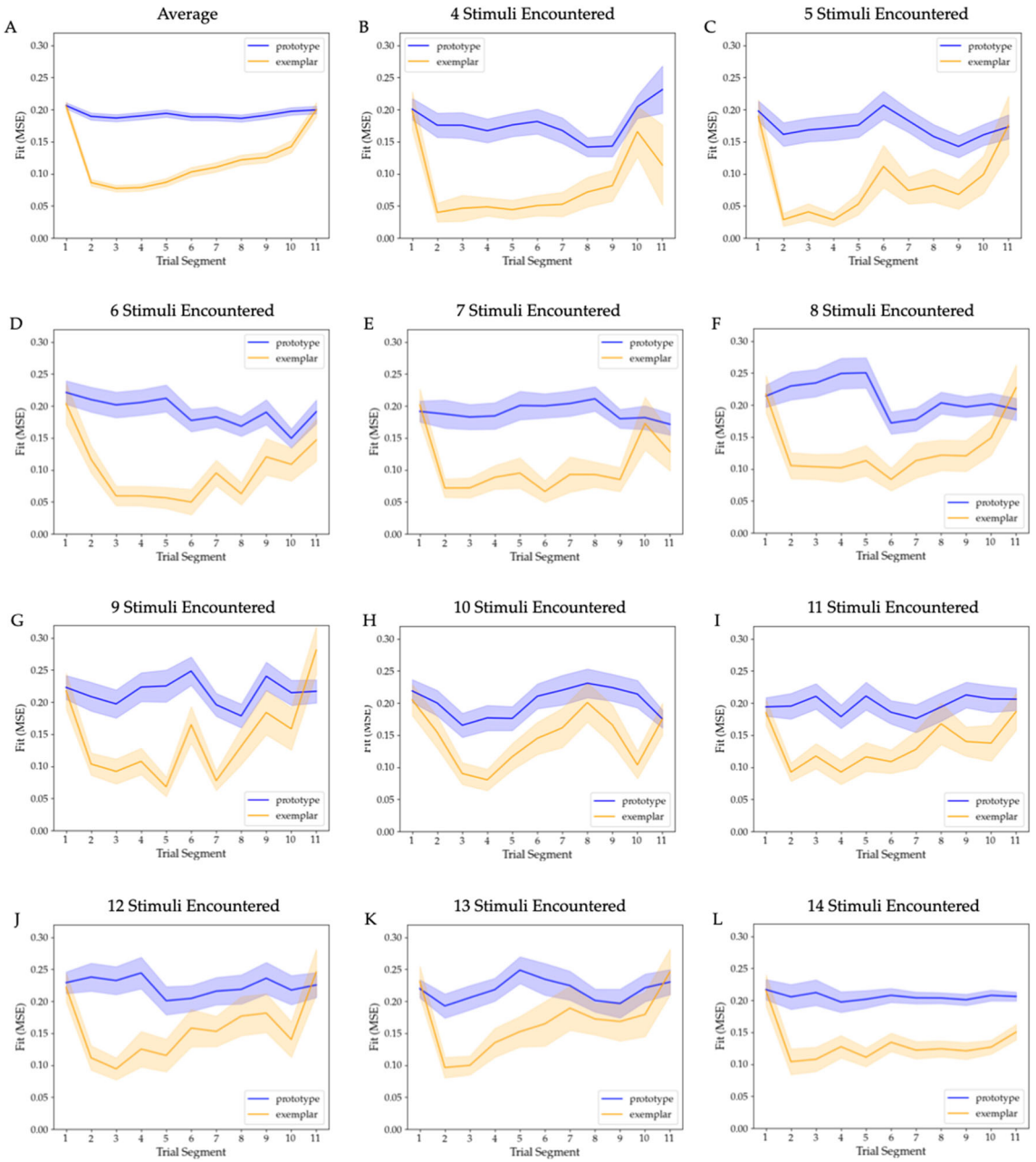

Fig. 14 Model fits for the experimental condition of Experiment 2 with four exceptions, where each subplot (**B**-**F**) only includes trials that have the specified, fixed number of stimuli seen thus far, i.e., **B** only graphs results from the sets of trials for each participant between the introductions of the fourth and fifth stimuli, and **A** averages these results, including the results from 1–3 stimuli seen (not shown individually)

number of exemplars is fixed and has no effect on the exemplar model's predictions. A formal permutation test on the averaged data from these analyses using the F interaction statistic to test the effects of model type (prototype/exemplar) and trial segment (second, last) was significant, $p < .001$. The results from this analysis in the experimental condition with four exceptions are shown in Fig. 14, which indicates that our previously observed results still hold after controlling for the number of exemplars: the exemplar model's advantage over the prototype model decreases over time in the experimental condition.

Open Practices Statement The data and analysis codes for all experiments are available on GitHub (https://github.com/arjundevraj/rational-categorization, https://github.com/arjundevraj/word_categorization). This study was pre-registered (Experiment 1: https://aspredicted.org/yq984.pdf; Experiment 2: https://aspredicted.org/9me7s.pdf).

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396–408.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37*, 372–400.

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition, 118*(1), 2–16.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short-and long-term recognition. *Psychological Science, 23*(6), 625–634.

Elliott, S. W., & Anderson, J. R. (1995). Effect of memory decay on predictions from changing categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 815.

Estes, W. K. (1994). *Classification and cognition*. Oxford University Press.

Griffiths, T., Canini, K., Sanborn, A., & Navarro, D. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 323–328).

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review, 95*(4), 528.

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.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527.

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*, 128–148.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 294–317.

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.

Medin, D. L. (1975). A theory of context in discrimination learning. In G. H. Bower (Ed.), *The psychology of learning and motivation.* Academic Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 241–253.

Medin, D. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 607–625.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*(5), 355.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*(1), 89–115.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 87–109.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700–708.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79.

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*, 924–940.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: learning, memory, and cognition, 14*(4), 700.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of experimental psychology: human perception and performance, 17*(1), 3.

Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology, 75*, 97–129.

Nosofsky, R. M., Cox, G., Cao, R., & Shiffrin, R. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology Learning, Memory, and cognition, 40*, 1524–39.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22*(3), 352–369.

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.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological review, 101*(1), 53–79.

Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new

recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1194.

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

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 548.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304–308.

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

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382–407.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology, 31*(2), 81–109.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General, 133*(4), 534–553.

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

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology, 24*, 1411–1436.

Zeng, T., Tompary, A., Schapiro, A. C., & Thompson-Schill, S. L. (2021). Tracking the relation between gist and item memory over the course of long-term memory consolidation. *eLife*, *10*, e65588.