Exemplar-Model Account of Categorization and Recognition

When Training Instances Never Repeat

Mingjia Hu and Robert M. Nosofsky

Indiana University Bloomington

Robert Nosofsky

Psychological and Brain Sciences

1101 E. Tenth Street

Indiana University

Bloomington, IN 47405

[nosofsky@indiana.edu](mailto:nosofsky@indiana.edu)

A classic debate in research on human category learning has involved the contrast between exemplar and prototype models. According to prototype models, people represent categories by abstracting their central tendencies from constituent training instances, and make categorization judgments based on similarity to the prototypes (Homa, Cross, et al., 1973; Reed, 1972; Smith & Minda, 1998). In contrast, according to exemplar models, people represent categories by storing individual constituent exemplars in memory, and base categorization judgments on similarity to the exemplars (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986).

Brief Review of the Debate in the Context of the Dot-Pattern Classification Paradigm

Many experimental results that initially appeared to strongly favor the prototype model were obtained from the classic dot-pattern paradigm introduced by Posner and Keele (1968); however, as reviewed below, exemplar theorists argued that such results are compatible with exemplar models. Recently, Homa, Blair, McClure, Medema, and Stone (2019) reported a new set of intriguing results derived from this paradigm that they claimed severely challenge exemplar models. Our goal in this article is to address these challenges, using both exemplar-based modeling approaches and new empirical studies. To set the stage for this goal, we first provide a brief review of the basic dot-pattern paradigm and its influence on the prototype-exemplar model debate.

In a typical dot-pattern experiment, prototypes representing different categories are first generated by randomly placing nine dots in a grid; then patterns of various levels of distortion are constructed by displacing the dots of the prototypes according to a statistical-distortion rule. Higher levels of distortion produce dot patterns that are systematically less similar to the originating prototypes. The experiment typically consists of a learning phase, in which subjects are trained to classify a number of distorted patterns, followed by a transfer phase, in which subjects are tested on classifying a variety of patterns including the old training distortions, the prototypes, and various new distortions of the prototypes. Early studies found that in the transfer phase, the classification accuracy of the prototype not presented in the learning phase was higher than that of the various new distortions, and sometimes even exceeded that of the old training distortions themselves (e.g., Homa et al., 1973; Posner & Keele, 1968, 1970). This prototype-enhancement effect was cited as evidence for the abstraction of a prototype as a basis for representing the categories. Moreover, classification accuracy tended to decrease for new patterns with higher levels of distortion from the prototypes, producing a systematic “typicality gradient”. The typicality gradient is also consistent with the prototype model as patterns of higher levels of distortion are less similar to the prototype.

However, these classic results were also shown to be compatible with the predictions from exemplar models (e.g., Hintzman, 1986; Nosofsky, 1988; Shin & Nosofsky, 1992), which posit that category evidence is related to the summed similarity of test items to the training exemplars of the categories. The typicality-gradient effect arises because the summed similarity of novel test items to the old exemplars of a category tends to be higher for patterns closer to the center of category. The prototype-enhancement effect arises because the prototype is highly similar to virtually all the old training distortions; by contrast, any given old training distortion may be highly similar only to itself.

Since the classic studies of Posner and Keele (1968, 1970), many prototype theorists have argued that the prototype-abstraction process in the dot-pattern paradigm is more likely to operate when category size (i.e. the number of distinct training exemplars) is large and/or when the transfer phase is delayed (e.g., Homa, Sterling & Trepel, 1981). For example, Homa et al. (1981) found that classification accuracy for a new distortion increased as a function of its similarity to specific high-level old training distortions; critically, however, the contribution of this specific new-old similarity effect to classification performance was attenuated as category size increased. It was also found that the old distortions were classified more accurately than the prototypes immediately after the learning phase, but that the reverse pattern was observed after a one-week delay. These effects were once considered strong evidence for a prototype-abstraction process. However, formal modeling of the specific new-old similarity × category size interaction and the differential forgetting of old distortions vs. prototypes revealed that both phenomena are qualitatively consistent with the predictions from pure exemplar models (e.g., Busemeyer, Dewey and Medin, 1984; Hintzman and Ludlam, 1980; Hintzman, 1986; Shin & Nosofsky, 1992). In general, the similarity of a new distortion to a specific old distortion makes a smaller relative contribution to overall summed similarity as category size increases: this specific new-old similarity tends to get “swamped” by similarity relations of the new distortions to the many other old training exemplars that compose the large-size categories. In addition, the differential forgetting of the old distortions versus the prototypes with delay of the transfer phase is also well-accounted for by pure exemplar models. The reason is that any given old distortion may be highly similar only to itself; thus, loss of fidelity in the memory representations for the old exemplars will be highly detrimental to old-item classification. By contrast, the prototype has high similarity to numerous old distortions that compose its category; this redundancy in the category representation protects the prototype from suffering major performance loss due to delayed testing.

More recently, Smith (2002) focused attention on the pattern of results observed in a particular version of the dot-pattern classification pattern introduced by Knowlton and Squire (1993). In this version, observers are exposed to 40 high distortions of a single prototype during an incidental training phase. Following the training, participants are tested on the prototype, new-low and new-high distortions of the prototype, and on random patterns not generated from the prototype. The participants are instructed to judge whether or not each test item is a member of the category that they experienced during the training phase. Smith (2002) argued that even though both exemplar and prototype models predict correctly the ordering of classification endorsements of the different pattern types (see Nosofsky & Zaki, 1998), the steepness of the typicality gradient observed in this paradigm falsifies exemplar models.

However, Zaki and Nosofsky (2004, 2007) provided clear evidence that the steep typicality gradient observed in this paradigm did not arise from the abstraction of a prototype from the training instances; instead, it was an artifact of the structure of the test phase used in this paradigm and a result of continued learning that took place during the test phase (for closely related findings and criticisms of Smith’s 2002 interpretations, see Palmeri & Flanery, 1999, 2002). Specifically, in the Knowlton-Squire (2003) paradigm, participants are flooded with numerous presentations of the prototype and its low distortions during the test phase. These test patterns are all centrally located in the category and are all highly similar to one another. As participants continually experience these patterns during the test phase, they continue to build upon the category representation that was developed during the incidental training phase by storing numerous new examples in the center of the category. Zaki and Nosofsky (2004, 2007) provided strong evidence in favor of this view by manipulating the structure of the test phase itself: they found dramatic changes in the shape and steepness of the typicality gradient as a function of these test-phase manipulations. Moreover, in all cases, a simple exemplar model provided excellent quantitative accounts of the shape and steepness of the typicality gradient that was observed across the different test-phase manipulations.

The New Challenge: Classification and Recognition When Exemplars Never Repeat

As briefly reviewed above, the prototype-exemplar debate in the context of the dot-pattern paradigm has a long history. However, the debate was recently renewed in an interesting new study reported by Homa, Blair, McClure, Medema, and Stone (2019), who claim to have reported results that pose substantial problems for exemplar models. The central purpose of the work reported in the present article was to address these new challenges.

In Homa et al.’s (2019) experiments, participants learned to classify dot patterns into three categories, and then engaged in various transfer tests.1 As in past versions of the paradigm, each individual category was generated around a dot-pattern prototype. Low, medium, and high distortions of each prototype were generated using the Posner-Keele (1968) statistical-distortion algorithm. Foil patterns were also used, which were medium distortions of prototypes that were not trained during category learning.

The key manipulation across the experiments involved the structure of the learning phase.

Two different learning phases were employed across two conditions. In both conditions, the learning phase was organized into a sequence of 15-trial blocks involving the presentation of medium-level distortions of the prototypes. In the *repeating* condition (REP), the same 15 medium-level distortions (5 per each of the 3 categories) were presented in every 15-trial learning block. By contrast, in the *non-repeating* condition (NREP), no individual training instance was ever repeated. Instead, 15 different medium-level distortions (5 per category) were presented in each 15-trial learning block. So, for example, in Homa et al.’s (2019) Experiment 1, in which there were 20 learning blocks, participants experienced a total of 15 distinct training instances in the REP condition (each one repeated 20 times), but experienced 300 unique training instances in the NREP condition (each one presented only one time).

Across experiments, participants then engaged in various transfer tests. In Experiment 1, participants were required to classify novel patterns (prototypes, low-, medium-, and high-level distortions) into the trained categories. In Experiments 2 and 3, participants instead engaged in old-new recognition tests, in which they judged whether test patterns had or had not been presented during the training phase. (In Experiment 2, the test patterns were old distortions, new-medium distortions, and foils; in Experiment 3, the test patterns were old distortions, new-medium distortions, and prototypes.)

Homa et al. (2019) focused on three main patterns of results in their study that they claimed severely challenged exemplar-only models, but that were well accommodated by a model that assumed exemplar-based classification in the REP condition, but prototype-based classification in the NREP condition. The first result was that, across the three experiments, there was no difference in speed of learning across the REP and NREP conditions. The second result was that, in the recognition-transfer tests, participants were unable to discriminate between the old- versus the new-medium distortions in the NREP condition, but showed well-above-chance discrimination of these pattern types in the REP condition. The third result was that participants classified test patterns with high accuracy in the transfer phase of the NREP condition, despite the fact that no single training instance was ever repeated during the learning phase, and despite the fact that in the recognition-transfer tests they showed no ability to discriminate between the old-medium and new-medium distortions in the NREP condition.

Plan of Current Research

The purpose of the present research was to further investigate and address the above-stated challenges, using both model-based approaches and testing of new experiments. To preview, in our view, the general pattern of classification and recognition results that Homa et al. (2019) reported in their transfer tests does not pose major qualitative challenges to exemplar models: Such models predict *a priori* that classification transfer to novel test items from the categories will be excellent in both the REP *and* the NREP conditions; that ability to distinguish old from new medium distortions in the recognition tests will be excellent in the REP condition; but that ability to distinguish old from new medium distortions in the recognition tests will likely be very poor in the NREP condition. As we discuss more fully later in our article, whether one observes old-new discrimination in the NREP condition that is significantly above chance will likely vary with individual-subject capabilities, detailed similarity relations among the patterns, and statistical-power considerations; but the essential point is that the exemplar model predicts that old-new discrimination performance will be poor in that condition, while at the same time predicting excellent classification of novel transfer items.

On the other hand, we acknowledge that Homa et al.’s finding across their three experiments that there was no difference in speed of learning across the REP and NREP conditions does indeed pose a fundamental challenge to exemplar models. One of the central purposes of our newly reported experiments was to pursue that intriguing result. We organize our presentation by first addressing the patterns of classification and recognition transfer data, and then turn to our new empirical investigations of the learning data.

Exemplar Model of Classification and Recognition Transfer in the REP and NREP Conditions

Our main approach to addressing the patterns of classification and recognition transfer data reported by Homa et al. (2019) is to report new model-based analyses of those data. However, before turning to the formal model-based analyses, we believe it is useful to develop an intuitive and conceptual account of the findings.

As explained in numerous previous articles (e.g., Nosofsky, 1988, 1991), according to the *generalized context model* (GCM), which is a well-known representative from the class of exemplar models, classification is based on a “relative-summed-similarity” rule. The evidence in favor of each category is found by summing the similarity of a test item to the training examples of each of the categories; if the summed similarity of the test item to the target category is relatively large, and its summed similarity to the training examples of the contrast categories is relatively small, then the test item will be accurately classified in the target category. By contrast, old-new recognition decisions are based on an “absolute-summed-similarity” rule: the evidence in favor of an old decision is found by summing the similarity of a test item to all the training examples of all the categories. Because different decision rules are involved, it is straightforward for the exemplar model to predict varieties of “dissociations” in which, say, classification performance is highly accurate, while recognition performance is extremely poor; and vice-versa (for numerous examples, see, e.g., Nosofsky, 1988, 1991).

From the perspective of the exemplar model, the basic scenario underlying the structure of Homa et al.’s (2019) REP and NREP conditions is illustrated schematically in Figure 1. In each condition, the medium-old distortions that serve as training exemplars (illustrated as x’s) form clouds around the category prototypes from which they were generated. In the REP condition, each individual training example is presented multiple times, so has a very strong memory representation (illustrated as boldface x’s); by contrast, in the NREP condition, each individual training example is presented only once, so has a weak memory representation. Note further that in the NREP condition, because so many individual training exemplars are generated from each prototype, the cloud that is produced will tend to be “denser” than in the REP condition; in other words, it will tend to provide better “coverage” of the multidimensional space in which the category patterns are embedded.

Now, suppose that an observer is tested with a novel transfer probe from one of the categories, say Category A (illustrated schematically as the lower-case red “o” in both the REP and NREP panels of the figure). Note that the probe is generated using the same statistical-distortion algorithm as is used to generate the training examples. Intuitively, it is easy to see that, regardless of whether an observer was trained in the REP condition or the NREP condition, the test probe will tend to have high relative similarity to the training examples of Category A, and low relative similarity to the training examples of Categories B and C. Thus, classification accuracy will tend to be high, regardless of training condition. The mere fact that no training example was ever repeated during the NREP condition has no bearing on that general prediction. (In the formal modeling analyses, we will address more fine-grained predictions involving the patterns of classification transfer accuracy reported by Homa et al.)

By contrast, according to the exemplar model, the predicted patterns of old-new recognition will depend crucially on whether observers were trained in the REP condition or the NREP condition. A fundamental component assumption of exemplar models is that the similarity between patterns decreases exponentially with their distance in psychological space (Shepard, 1987). An item is maximally similar to itself, and the similarity tends to drop off rapidly as the distance between patterns increases. Therefore, in the REP condition, old-new recognition discrimination between old- and new-medium distortions will tend to be high, for two reasons. First, in computing absolute summed similarity, the maximal self-match of an old test item to its representation in memory is multiplied N times, where N is the number of repetitions of the item during training. This multiplied self-match contribution dominates the absolute summed-similarity computation in the REP condition, providing old test items with a strong recognition signal. The recognition signal is much weaker for the new test items, because no self-match contribution to the absolute summed-similarity computation is present. Second, because the cloud of training examples tends to be “sparse” in the REP condition, many of the new distortions that are tested in that condition will not have any close old-training-example neighbors, leading to an even lower absolute-summed-similarity signal.

The situation is quite different in the NREP condition, as illustrated schematically in the bottom panel of Figure 1. Here, each training example has been presented only once. Thus, when tested with an old training item, there is only a single self-match contribution made to the absolute-summed-similarity signal. The relative contribution of the single self-match is “swamped” by the similarity of the test item to all the other items presented during training. Furthermore, because the cloud of training examples is “dense” in the NREP condition, it is highly likely that each tested new medium distortion will be highly similar to at least some of the old training examples. A consequence is that there may be virtually no difference between the absolute-summed-similarity signals associated with the old- and new-medium distortions in Homa et al.’s (2019) NREP condition. Thus, old-new recognition discrimination would tend to be quite poor in that condition.

In the remainder of this section, we attempt to go beyond the intuitions offered above by developing a formal simulation-based modeling account of the detailed classification and recognition transfer data reported by Homa et al. (2019) in both the REP and NREP conditions.

.

Simulation-Based Exemplar-Modeling Account

Although Homa et al. (2019) presented certain formal-modeling accounts of their data in their original article, there were some limitations in their modeling approach that we address throughout our article (see below). In general, our view is that formal modeling of data in the classic dot-pattern paradigm presents some significant challenges, mainly because the psychological dimensions that compose the patterns are unknown. In some approaches, researchers use the physical coordinates of the 9 dots in each pattern as a representational scheme, and compute similarities between patterns based on distances between corresponding dots in each of the patterns (e.g., Palmeri & Flanery, 2002; Smith, 2002; Zaki & Nosofsky, 2007). There are several limitations of this approach. First, past research has made clear that the configurations of dots in the patterns give rise to salient emergent dimensions that are not captured by the physical dot locations themselves (Hock, Tromley, & Polmann, 1988). For example, the approach fails to capture emergent properties such as overall shape of the patterns, symmetry, orientation of the configurations, geometric sub-parts, and so forth. Second, even if one resorts to computing the physical dot-location distances, there is the problem of how to establish correspondences between the dots of the patterns (i.e., which dots in pattern 1 should be lined up with which dots in pattern 2 to compute their physical distance?). This question is especially problematic in cases in which one computes distances between dot patterns generated from different prototypes, or between category patterns and random foils (for extensive discussion of this difficult problem, see Palmeri & Flanery, 2002). Ultimately, such decisions involving between-pattern dot correspondences are arbitrary, making the physical dot-distance approach a highly questionable one.

In an alternative approach, Shin and Nosofsky (1992) and Palmeri and Nosofsky (2001) collected similarity judgments among pairs of dot patterns, and used multidimensional scaling techniques to locate the patterns as points in psychological spaces. This approach is not practical, however, in cases involving large sets of patterns (i.e., the case in the present kinds of experiments), because the number of similarity judgments required for the MDS analysis becomes astronomical.

In still another approach, researchers simply define free parameters representing the average similarity among different types of patterns, and substitute these parameter estimates into exemplar-model and prototype-model equations for predicting classification and recognition (e.g. Busemeyer et al., 1984; Homa et al., 1981). Indeed, this approach is the one that Homa et al. (2019) followed in their own modeling analyses. Although the approach is a reasonable one, it too has various limitations. First, it often involves a proliferation of free-parameter estimation with insufficient constraints on the settings of the parameters. Second, it fails to capture the critical fact that there are enormous individual-stimulus differences across different tokens of the types of patterns. For example, based on their random positioning in psychological space that arises from use of the statistical-distortion algorithm, some medium-level distortions may be extremely easy to classify and/or recognize, whereas others may be extraordinarily difficult. Estimating an “average similarity” parameter across different types of patterns fails to capture this form of individual-stimulus variability and complexity.

Until these challenges are solved, our view is that rigorous quantitative comparisons among competing models in this dot-pattern domain are probably not advisable. Instead, in the approach that we take here, our goal is to demonstrate that exemplar models provide viable qualitative accounts of the major patterns of results of interest that were reported by Homa et al. (2019). Furthermore, following Hintzman’s (1986) influential style of modeling, we adopt a simulation approach in which patterns are constructed that are analogous to the intended psychological structure of the dot-pattern stimuli and categories, without making any claims of direct correspondence.

In our simulation approach, we presume that the patterns occupy points in a six-dimensional psychological space. (The outcomes of the simulation analyses are not greatly affected by the number of chosen dimensions; we chose to use six because previous MDS work reported by Shin and Nosofsky [1992] revealed that 6-dimensional solutions provided good accounts of similarity relations among the types of dot-patterns used in these studies.) For each individual simulation and for each category, a prototype is generated by randomly choosing values in the range (0, *between*) along each of the six dimensions, where *between* is a freely estimated parameter. In general, larger values of *between* will produce category prototypes with larger distances from one another, resulting in greater levels of between-category dissimilarity.

Next, statistical distortions of each prototype are produced by randomly sampling *z* scores from a standard normal distribution, and adding scaled values of those z scores to the prototype dimension values. Specifically, let *Pim* denote the value of Prototype *i* on dimension *m*. For simplicity and to reduce the number of free parameters, we define values *low*, *medium*, and *high* as average dot-distance movements produced by the Posner-Keele statistical-distortion algorithm for these types of patterns. Homa et al. (2019, p. 398) report these values to be *low* = 1.20, *medium* = 2.80, and *high*=4.60. We define a free scaling parameter *within*, which will primarily influence the degree of within-category dissimilarity among the patterns in each category in our simulations. Let *xim* denote the value of a statistical distortion produced from Prototype *i* on dimension *m*. The statistical distortions along each dimension in our simulations for low, medium, and high distortions were produced as follows:

*xim* = *Pim* + *within*\**low*\**z*, for low distortions

*xim* = *Pim* + *within*\**medium*\**z*, for medium distortions

*xim* = *Pim* + *within*\**high*\**z*, for high distortions

(Of course, in generating the distortions, a new random value of *z* is sampled for each distortion along each individual dimension. Also, note that analogous to Homa et al.’s experimental method, “foil” patterns were produced using the above algorithm by creating medium-level distortions from randomly generated prototypes that were not used to generate training instances during the learning phase.)

Use of this simulation algorithm was intended to produce prototypes and clouds of statistical distortions around those prototypes analogous to the schematic illustrations in our Figure 1.

Once the patterns are created for each individual simulation, standard equations from Nosofsky’s (1986, 1988, 2011) exemplar model are used to generate predictions of classification and recognition in the transfer phases of Homa et al.’s experiments. In particular, the standard Euclidean distance formula is used to compute the distance between test-item *i* and training-example *j*,

(1)

The similarity between test-item *i* and example *j* (*sij*) is an exponential-decay function of this distance,

(2)

where *c* is a sensitivity parameter that describes the rate at which similarity declines with distance. The sensitivity parameter provides a measure of overall discriminability among patterns in the psychological space.

The probability that a test-pattern *i* from, say, Category A, is correctly classified in Category A, is then found by summing its similarity to all the training examples *a* that belong to Category A, and dividing by the summed similarity of *i* to all the training examples of all the categories:

(3)

Equation 3 is the “relative-summed-similarity” rule for classification that we described in intuitive terms earlier in this section. In Equation 3, the parameter γ is a response-scaling parameter. When γ = 1, the observer responds by “probability-matching” to the relative summed similarities of each of the categories; as γ grows greater than 1, the observer responds more deterministically with the category that yields the largest relative summed similarity (for extensive discussion as well as process-model interpretations for the emergence of the γ parameter, see, e.g., Ashby & Maddox, 1993; Nosofsky & Palmeri, 1997; Nosofsky et al., 2002).

Although not made explicit in the notation, note that in applying Equation 3, the similarity to each individual training example is being summed *N* times in the REP condition, where *N* is the number of training blocks (i.e., the number of times the specific example is repeated during training). By contrast, the similarity to each individual training example enters only once in the NREP condition, because no individual training example was ever repeated.

Finally, according to the exemplar model, the probability that test-item *i* is judged to be “old” in the recognition-transfer tests is found by summing the similarity of the test item to all the examples of all the categories, and entering that absolute summed similarity into the following choice rule:

(4)

(Again, although not made explicit in the notation, the similarity to each individual training example is being summed *N* times in the REP condition, but only once in the NREP condition.)

In Equation 4, the parameter *k* is a criterion parameter that influences the overall bias for making old versus new judgments. Observers presumably adjust the setting of *k* in accord with the overall levels of absolute summed similarity being generated by test patterns presented during the test phase (for extensive discussion, see, e.g., Nosofsky et al., 2011). (For example, if very large values of absolute summed similarity are being generated for test patterns in a condition, then the observer will presumably adopt a large setting of *k*; whereas if small values of absolute summed similarity are being generated, then the observer will adopt a small setting of *k*.) Because the absolute-summed similarity levels differ significantly across the REP and NREP conditions and across the different recognition-transfer conditions tested in Homa et al.’s Experiments 2 and 3, separate criterion parameters are estimated for each of these separate conditions.

Fits of the Model to the Classification and Recognition Transfer Data

The classification- and recognition-transfer probabilities observed by Homa et al. for the different item types across their Experiments 1-3 are reproduced here as colored bars in the panels of Figure 2. As described earlier, classification accuracy was high for all the pattern types in both the REP and NEP conditions (see Figure 2, panel A). In addition, Homa et al. observed the classic “typicality gradient” across both conditions, in which classification accuracy was highest for the prototypes, followed in order by the low-, medium-, and high distortions. Classification accuracy during the transfer phase tended to be slightly higher in the NREP condition than in the REP condition (although the difference was statistically significant only for the medium distortions).

In the Experiment-2 recognition-transfer phase (see Figure 2, panel B), in the REP condition, participants judged old-medium distortions to be old with significantly higher probability than they judged new-medium distortions to be old; however, there was no difference in old-recognition probabilities for the old- versus the new-medium distortions in the NREP condition. In both conditions, the foils received the lowest old-recognition probabilities. The same pattern of recognition probabilities for the old- and new-medium distortions was observed in the REP and NREP conditions of Experiment 3 (see Figure 2, panel C); in addition, endorsements of the prototypes as old were essentially the same as for the old distortions in the REP condition, but significantly exceeded the endorsement probabilities for the old distortions in the NREP condition.

We fitted the simulation-based exemplar model to these data by searching for the free parameters in the model that minimized the sum-of-squared deviations between the predicted and observed probabilities for each of the item types across all the conditions. As described earlier, the free parameters included the between-category dissimilarity parameter *between*; the within-category dissimilarity parameter *within*; the sensitivity parameter *c*; the response-scaling parameter γ; and the settings of the response-criterion parameter *k*. The parameters *between*, *within*, *c*, and γ were held fixed across all experiments and conditions. Separate values of the response-criterion parameter *k* were estimated for each of the REP and NREP conditions across Experiments 2 and 3. We conducted 10,000 simulations in generating the predictions, and used the Hook and Jeeves (1968) parameter-search algorithm to locate the best-fitting parameters.

The predictions from the exemplar model are shown as solid dots in Figure 2, with best-fitting parameters reported in Table 1. Although our goal involved achieving only a reasonable qualitative account of the pattern of results, it turns out that the quantitative fit to the data is nothing short of outstanding. All of the major qualitative patterns described above for both the classification and recognition data are captured by the model, and usually with high quantitative precision.

It can be seen from inspection of Figure 2 that the exemplar model predicts a substantial difference in old-recognition probabilities for the old- versus the new-medium distortions in the REP conditions of Experiments 2 and 3, but predicts a minuscule difference for these pattern types in the NREP condition. As discussed earlier, one of the major results that Homa et al. (2019) emphasized in their study was that participants showed zero ability to distinguish between the old- versus the new-medium distortions in the NREP condition of their recognition-transfer tests. As can be seen from our model-based predictions, tracking down the predicted minuscule difference would likely require an enormous amount of data collection. In addition, in our General Discussion, we will express some other concerns about the methods used in Homa et al.’s (2019) experiments that may also have made it difficult to detect any difference in recognition probabilities for these pattern types.

Beyond accounting for the patterns of recognition for the old- and new-medium distortions across the REP and NREP conditions, it is also of interest to note that the exemplar model provides an excellent account of the high recognition-endorsement rates for the prototypes in Homa et al.’s Experiment 3. As explained earlier, although the prototype was never presented during training, it has high similarity to numerous of the old training distortions. Thus, its absolute summed similarity is high, even exceeding that of the individual old distortions in the NREP condition.

Inspection of the best-fitting parameters (Table 1) reveals, as one would expect, that the between-category distance estimate greatly exceeds the within-category distance estimate. (Further analysis revealed that essentially the same fits are achieved as long as the ratio of these parameter estimates is held roughly constant, so the model is achieving its good fits with even fewer effective parameters than reported here.) In addition, the magnitude of the recognition-criterion parameter *k* varies across the conditions in sensible ways. In general, absolute-summed similarity tends to be greater in the REP condition than in the NREP condition, so participants set a stricter criterion *k* in REP than in NREP. Likewise, because the prototypes are substituted for the foils across Experiments 2 and 3, average absolute summed similarity is slightly higher in Experiment 3 than in Experiment 2, and there is a slight adjustment in the magnitude of *k* consistent with this change. We defer discussion of the *c* and γ parameter estimates until after presentation of the results from our new experiments.

The Learning Data

As we argued at the outset of our article, the real challenge to exemplar-model predictions from Homa et al.’s (2019) findings does not lie in the patterns of classification and recognition-transfer data; instead, it lies in the learning data. Across their three experiments, Homa et al. observed no significant differences in the speed of classification learning across the REP and NREP conditions. Indeed, the learning curves across these conditions either lied virtually on top of one another (see Homa et al., 2019, Figures 1 and 3), or there tended to be a slight advantage for NREP (see Homa et al., 2019, Figure 2).

The exemplar model has been formalized mainly to account for patterns of performance at time of transfer, and the details of the complicated early learning processes that are involved have been left for future research.2 Nevertheless, we agree with Homa et al. (2019) that formalizations of the exemplar model intended to capture the main qualitative learning effects would certainly predict that speed of learning should be faster in the REP condition than in the NREP condition. The basic idea, as Homa et al. explained in detail in their original article, is that the repeated training examples presented in the REP condition make maximal contact with their own representations in memory, so summed similarity to the correct target category grows more rapidly in the REP condition than in the NREP condition.

To illustrate, following Nosofsky and Kruschke (1992) and Stanton and Nosofsky (2013), a rudimentary learning version of the exemplar model can be formalized by extending the Equation-3 classification rule with a “background-noise” constant β:

(5)

As learning proceeds, the summed-similarity terms in the equation grow, because one is summing similarities to larger and larger collections of stored exemplars. Early in learning, the summed-similarity terms are small in magnitude, and the background constant β dominates: here, the model predicts responding that is close to chance. As learning proceeds and the summed-similarity terms grow larger, the influence of the background-noise constant fades away, and responding is governed by similarity comparisons to the stored examples.

In Figure 3 we show predicted learning results from the simulation-based version of the exemplar model, with parameters held fixed at those values used to fit the transfer data (see Table 1), but with the model extended with the background-noise constant as in Equation 5. Clearly, the qualitative prediction from the exemplar model is that learning should proceed more rapidly in the REP condition than in the NREP condition, and this prediction is strongly disconfirmed by Homa et al.’s reported data.

We found this pattern of learning results reported by Homa et al to be so intriguing that our first inclination was to consider extensions of the exemplar model that might explain the null learning effects of REP versus NREP reported by these researchers. Before launching into this new theoretical investigation, however, we decided to repeat their basic experiment in order to achieve greater insights about the learning processes that might be operating. This goal served as the main motivation for the new experiments that we now report.

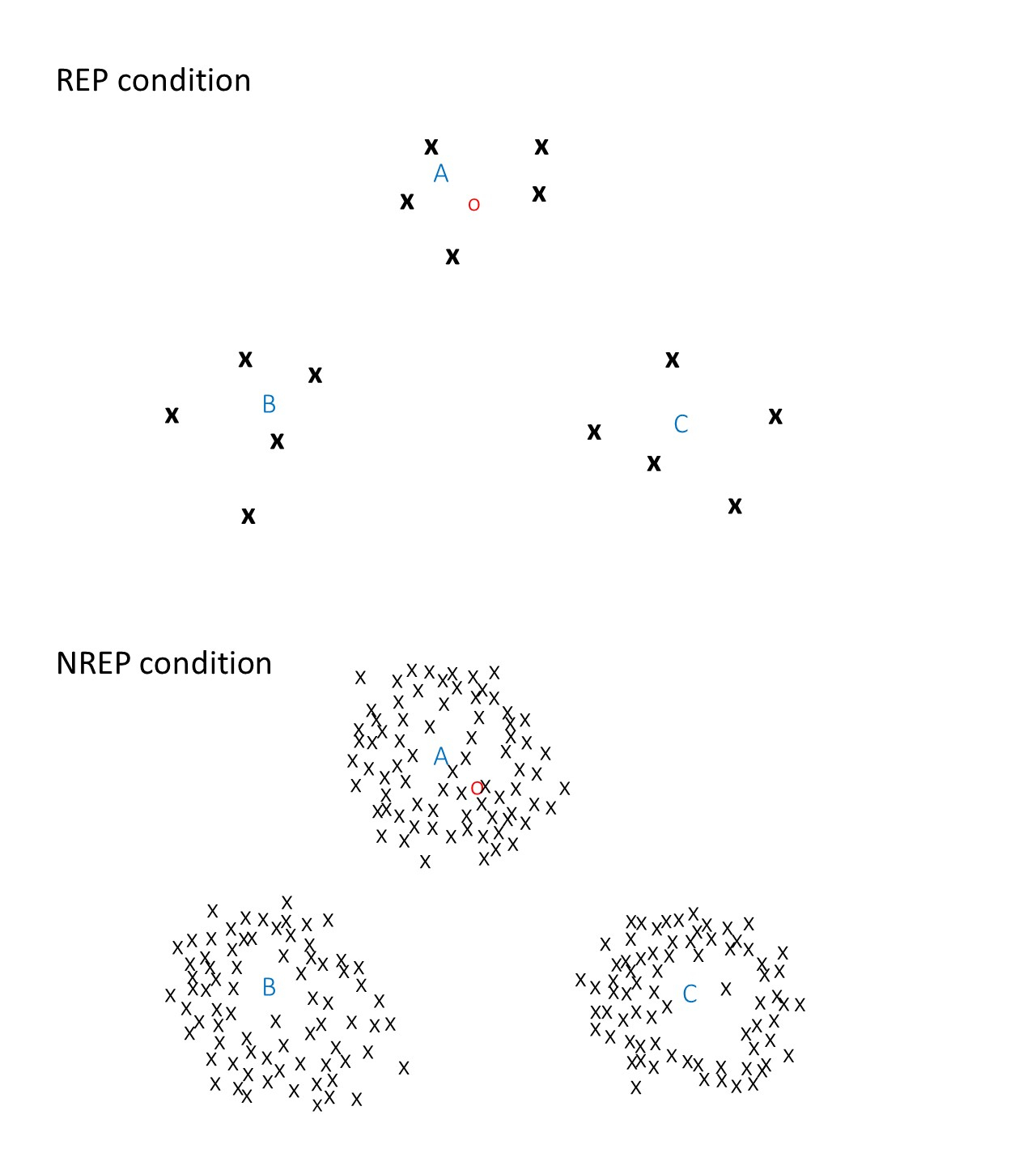


Figure 1. Schematic illustration of similarity structures of stored exemplars in the REP (top panel) and NREP (bottom panel) conditions. The letters A, B and C denote the prototypes for three different categories. Each symbol x denotes an old medium-level distortion in any category (boldfaced in the REP condition to indicate stronger memory traces due to their repeated presentations).  The red o in each panel denotes a new medium-level distortion in category A.

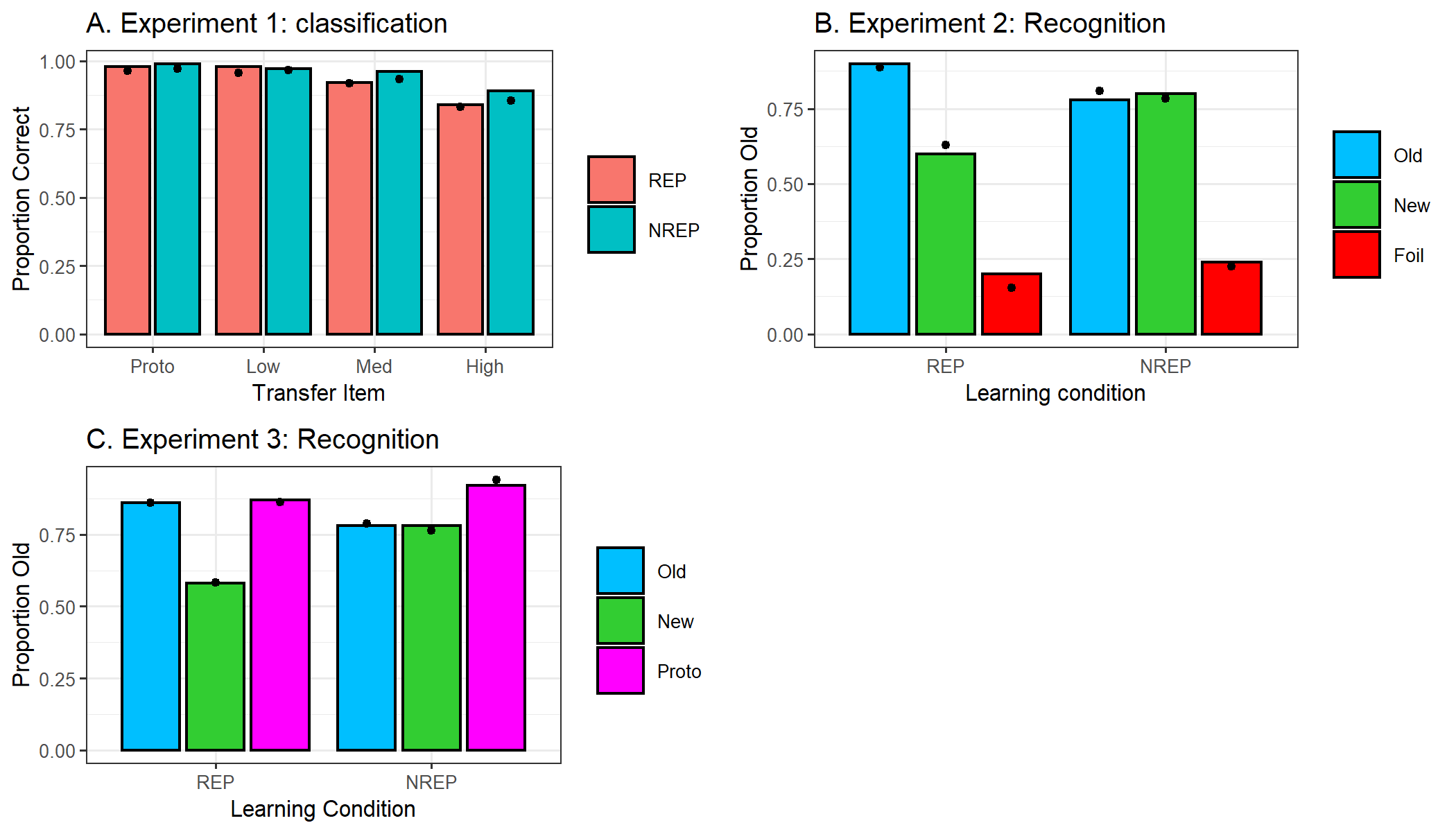


Figure 2. Observed (colored bars) and predicted (black dots) probabilities of correct classification and old recognition judgments in Homa et al.’s (2019) Experiments 1-3 (three-category conditions). Panel A: Experiment 1. Classification accuracy for the different types of transfer items in the REP and NREP conditions. Panel B: Experiment 2. Proportion of old-recognition responses for the different item types in the REP and NREP conditions. Panel C: Experiment 3. Proportion of old-recognition responses for the different item types in the REP and NREP conditions. Predictions are from the simulation-based version of the exemplar model described in the text.

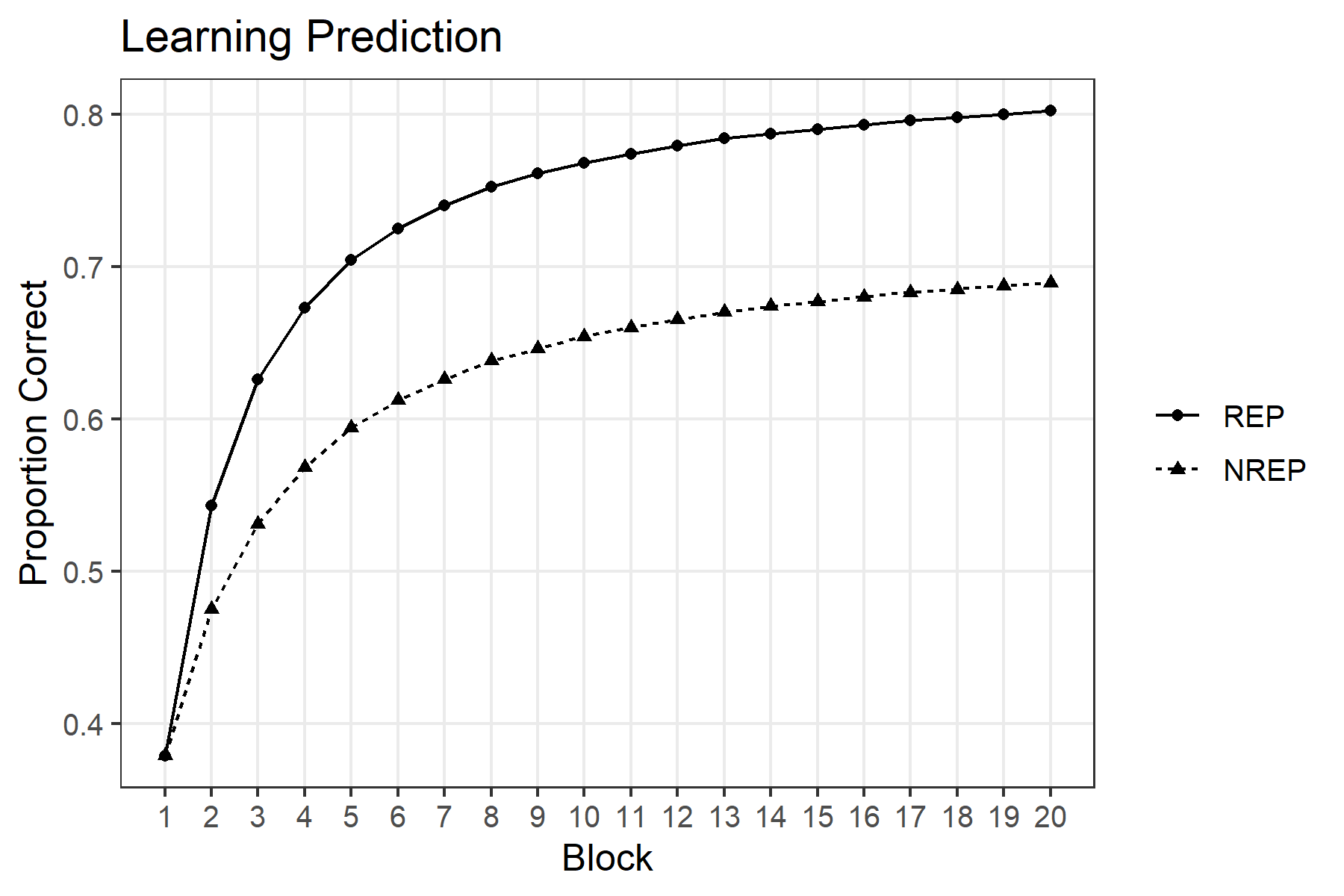


Figure 3. Predicted learning rates for the REP and NREP conditions from the version of the exemplar model extended with the background-noise constant β (Equation 5 in the text). Here, the background-noise constant is set at β=2, with all other parameters held fixed at the values listed in Table 1. The same qualitative pattern of results is predicted across an extremely wide range of the settings of the various free parameters.

Footnotes

1. In some experiments, two-category conditions were also tested, which yielded the same pattern of results as the three-category conditions.

2. Kruschke (1992) developed an influential connectionist-learning version of the exemplar model; however, acknowledged limitations of that model led him to theorize about a number of other learning mechanisms, and the search for a more satisfactory detailed learning model is still a work in progress.

Homa, D., Blair, M., McClure, S. M., Medema, J., & Stone, G. (2019). Learning concepts when instances never repeat. *Memory & cognition*, *47*(3), 395-411.

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

Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. *Psychology of learning and motivation*, *41*, 141-190.

Posner, M. I., Goldsmith, R., & Welton Jr, K. E. (1967). Perceived distance and the classification of distorted patterns. *Journal of experimental psychology*, *73*(1), 28.

Hock, H. S., Tromley, C., & Polmann, L. (1988). Perceptual units in

the acquisition of visual categories. *Journal of Experimental Psychology:*

*Learning, Memory, & Cognition*, **14**, 75-84.

Palmeri, T.J., & Nosofsky, R.M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. Quarterly Journal of Experimental Psychology, 54A(1), 197-235.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 638.