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 (1993) 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 exemplar-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 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 sets of 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 (Panels B and C) 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 very small 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 small predicted difference with adequate power 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, so 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. (In this example, the background-noise constant is set at β=2. The same pattern of predictions holds across an extremely wide range of the settings of the free parameters.) 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 the 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.

Experiment 1

The purpose of Experiment 1 was to replicate the learning and recognition-transfer phases conducted in Homa et al.’s (2019) Experiments 2 and 3. The structure of the learning phases in the REP and NREP conditions was the same as in Homa et al.’s experiments. Whereas Homa et al. had separated the testing of the foil and prototype patterns across their Experiments 2 and 3, we instead conducted a single experiment in which both the foils and prototypes were tested within a single transfer phase. (Of course, we continued to test the old-medium distortions and the new-medium distortions as well.) By testing the foils and prototypes within the same transfer phase, we introduced stronger constraints for modeling, because a common criterion setting is now required for predicting the false-alarm rates associated with both pattern types.

We expected the main pattern of recognition-transfer data to be roughly consistent with Homa et al.’s results, although we tested a larger sample size of participants in order to increase the possibility of detecting any small difference in recognition probabilities between the old and new medium distortions in the NREP condition. The critical question was whether or not we would replicate Homa et al.’s finding of no difference in speed of learning across the REP and NREP conditions.

Method

The study was approved by the Indiana University Institutional Review Board.

Subjects

The subjects were 198 undergraduates from Indiana University who participated in partial fulfillment of an introductory psychology course requirement. There were 98 subjects in the repeating (REP) condition and 100 subjects in the non-repeating (NREP) condition. Subjects were randomly assigned to the conditions. All subjects had normal or corrected-to-normal vision.

Stimuli and apparatus

The stimuli used in this experiment were dot patterns generated using Posner, Goldsmith, and Welton's (1967) procedure (i.e., the algorithm incorporated by Posner and Keele, 1968, 1970). Each prototype consisted of 9 dots randomly positioned in the central 30 × 30 area of a 50 ×50 grid and connected with white lines. For each individual subject, prototypes for six different categories were randomly generated. Three of the prototypes were used to generate training and transfer patterns for each of three categories; the remaining three were used to generate foils for the recognition-transfer phase.

Different training and transfer patterns of each category were generated using the statistical-distortion procedure of Posner et al. (1967). Each pattern was constructed from the prototype of its category by displacing each dot by a random distance and direction in accord with the Posner et al. procedure. Low-level, medium-level and high-level distortions were produced by displacing the individual dots, on average, 4, 6 and 7.7 Posner-levels away from their prototype. The foils used in the transfer phase were medium-level distortions of three randomly generated prototypes that were not used to generate category-training patterns.

Each individual subject was presented with a unique set of randomly generated prototypes and training and transfer patterns, with the only constraint being that the patterns were generated using the Posner et al. (1967) procedure.

We used Dell Computers to display the stimuli and control the experiment. The patterns were white in color and displayed at the center of a grey computer screen.

Procedure

In both the REP and NREP conditions, a standard learning-transfer paradigm was used. In the learning phase, subjects were instructed to classify dot patterns into three categories A, B and C. On each trial a pattern was presented on the screen and the subject classified it into one of the categories by pressing a corresponding button on the computer keyboard. Following the response, the computer provided immediate feedback informing the subject of the correct category. All patterns presented during the learning phase were medium-level distortions of the prototypes. In both the REP and NREP conditions, the learning phase consisted of 15 blocks, each of which had 15 trials (225 trials total).

In the repeating (REP) condition, there were 5 unique learning patterns for each of the three categories (15 learning patterns total). The same 15 learning patterns were repeated across the 15 blocks with the order of presentation randomized within each block. In the no-repeating (NREP) condition, there were 75 unique learning patterns for each category. Within each block, 5 unique learning patterns from each category were presented in a random order. No single learning pattern was ever repeated during the learning phase.

Following the learning phase, there was a recognition-transfer phase. On each trial, a single pattern was presented and subjects were instructed to recognize whether the pattern was old (presented in the learning phase) or new (not presented in the learning phase) by pressing a labeled button on the computer keyboard (J=old, F=new). No corrective feedback was provided on any trial.

In both the REP and NREP conditions, the transfer patterns consisted of 15 old distortions that were presented in the learning phase, 3 prototypes (1 per category), 15 new medium-level distortions (5 per category), and 6 foils (2 medium-level distortions generated from each of 3 prototypes not used to generate patterns in the learning phase). Each pattern was presented once in a random order for each subject for a total of 39 trials. In the REP condition, the 15 old distortions were the 15 unique patterns presented during the learning phase. In the NREP condition, the 15 old distortions were randomly sampled from the 225 learning patterns, with the constraints that no two patterns had been presented in the same learning block and that an equal number of patterns from each category was presented.

In both the learning and transfer phases, each pattern was presented centered on the computer screen and remained visible until a subject responded with a key press. In the learning phase, the corrective feedback on each trial appeared for 0.5s below the presented pattern. All subjects were tested individually in private, sound-attenuated cubicles.

Results

Prior to conducting detailed statistical and modeling analyses, we conducted preliminary analyses to identify severe outlier subjects within each condition. In the learning phase, we computed mean proportion correct for each subject during the final 8 blocks. In the transfer phase, we computed the difference between mean proportion of old judgments on the old learning patterns and the foils. We deleted from all subsequently reported analyses the data of any subject who performed more than 2.5 standard deviations below the mean on either measure. We deleted 7 subjects from the REP condition (leaving 91 valid subjects) and 5 subjects from the NREP condition (leaving 95 valid subjects). The main patterns of results from all subsequently reported statistical and modeling analyses were essentially the same if all subjects were included in the analyses.

Learning

The proportions of correct responses across the 15 blocks in the learning phase for the REP and NREP conditions are shown in the top panel of Figure 4. As can be seen, performance improved considerably across the learning blocks. More important, following the very early blocks, learning performance in the REP condition was considerably better than in the NREP condition. To confirm these observations, we conducted a 2x15 mixed-model ANOVA using learning condition (REP vs. NEP) and blocks as factors. The analysis3 revealed a significant main effect of blocks, F(8.66,1593.89) = 140.37, p < .001, η2 = .433, MSe = 3.427. The main effect of learning conditions was also significant, F(1,184) = 16.26 , p < .001, η2 = .081, MSe = 4.049, as was the interaction effect between learning condition and blocks, F(8.66,1593.89) = 2.463 , p = .01, η2 = .013, MSe = 0.606.

Transfer-Recognition

The probability with which each type of transfer pattern was judged as old in the REP and NREP conditions is shown in the top panel (A) of Figure 5. As expected, in the REP condition, old-recognition probability for the old medium-distortion learning patterns (*M*=.845) was considerably greater than for the new medium distortions (*M*=.343), and was also somewhat greater than old-recognition probability for the prototypes (*M*=.784). By contrast, in the NREP condition, old-recognition judgments were greatest for the prototypes (*M*=.916). Interestingly, however, even in the NREP condition, old-recognition probability was greater for the old medium-distortion learning patterns (*M*=.693) than for the new medium distortions (*M*=.632). Recognition probabilities for the foils were by far the lowest in both the REP (*M*=.053) and NREP (*M*=.151) conditions.

To confirm these observations, we conducted a 2x4 mixed-model ANOVA, using condition (REP vs. NREP) and item type (old, new-medium, prototype, foil) as factors. The analysis revealed a significant main effect of item type, F(2.67,490.68) = 883.93, p < .001, MSe = 23.835; a significant main effect of learning condition, F(1,184) = 54.85, p < .001, MSe = 1.565; and a significant interaction between the two factors, F(2.67,490.68) = 64.66, p < .001, MSe = 1.744. In the REP condition, the old-recognition probability for the old distortions was significantly greater than for the new medium distortions, t(90) = 24.51, p <.001, Cohen’s d = 2.569; and the increased recognition probability for the old distortions compared to the prototype was marginally significant, t(90) = 2.21, p = .059.4 Although the difference was much smaller than in the REP condition, even in the NREP condition the old distortions were judged as old significantly more often than the new medium distortions, t(94) = 3.59, p = .001, Cohen’s d = .368. However, in the NREP condition, the prototypes were judged as old with significantly greater probability than were the old distortions, t(94) = 10.21, p < .001.

Discussion

Contrary to Homa et al.’s (2019) results, we found that the speed of learning was significantly faster in the REP condition than in the NREP condition, implying that speed of category learning was indeed facilitated when the training patterns were repeated in each learning block. Furthermore, the magnitude of the advantage was substantial, averaging 0.095 across the final eight blocks. As explained earlier, this pattern is as predicted by exemplar models of classification learning. We return to a fuller discussion of the finding in our General Discussion after presenting the results from our Experiment 2.

Consistent with Homa et al.’s findings, the recognition-transfer data showed that subjects easily discriminated old medium-distortion and new medium-distortion patterns in the REP condition but had difficulty discriminating these patterns in the NREP condition. Nevertheless, even in the NREP condition, recognition probabilities for the old medium-distortions were significantly greater than for the new medium-distortions. As explained and demonstrated through simulation modeling in our introduction, this pattern of recognition-transfer effects is as predicted by exemplar models. As we discuss more fully in the General Discussion, whether the small-size recognition advantage that is predicted for old- compared to new-medium distortions in the NREP condition reaches statistical significance will undoubtedly vary with factors such as the ability and motivation levels of the participating subjects, precise similarity relations of the tested patterns to the training patterns, and statistical-power considerations.

Finally, we closely replicated Homa et al.’s findings involving the false alarm rates of the prototypes and foils: the prototypes were almost as likely to be judged as old as were the old-medium distortions in the REP condition, and were even more likely to be judged as old as were the old-medium distortions in the NREP condition. The high false-alarm rates of the prototypes in this paradigm are generally consistent with the qualitative predictions from exemplar models, because the prototypes have high similarity to numerous old training examples stored in memory; we test the adequacy of our simulation-based exemplar model to account for these prototype effects in our subsequent Modeling section. Not surprisingly, the false alarm rates of the foils were quite low regardless of the learning conditions.

Experiment 2

As we will argue more fully in our General Discussion, our finding in Experiment 1 that speed of learning was faster in the REP condition than in the NREP condition seems an intuitively sensible result. In our view, it is Homa et al.’s null-effect finding of no speed-of-learning differences that is the surprising one. Nevertheless, given the dramatic contrast in findings across Homa et al.’s experiments and ours, we decided to repeat the learning phase of our Experiment 1 in a new Experiment 2 with a new group of participants to test for the reliability of our findings.

A second purpose of Experiment 2 was to collect classification transfer data rather than recognition transfer data (replicating Homa et al.’s Experiment 1). We expected to replicate Homa et al.’s finding of the classic “typicality gradient” across both the REP and NREP conditions, with classification accuracy being highest for the prototypes, followed in order by the new-low, new-medium, and new-high distortions (a pattern that we have already shown is consistent with the predictions from the exemplar model). The main purpose of collecting the classification-transfer data was to provide additional constraints for model fitting: Our goal is to test the exemplar model on its ability to account jointly for the classification and recognition transfer data collected across our Experiments 1 and 2 in both the REP and NREP conditions. A minor variation from Homa et al.’s Experiment-1 procedure is that we also included tests of the old training distortions as part of the classification-transfer tests, to provide still further constraints for the formal modeling.

Method

Subjects

The subjects were 89 undergraduates from Indiana University who participated in partial fulfillment of an introductory psychology course requirement. There were 43 subjects in the REP condition and 46 subjects in the NREP condition. Subjects were randomly assigned to the conditions. All subjects had normal or corrected-to-normal vision. Although we had intended to collect a larger sample size, roughly matching the sample size from Experiment 1, the COVID-19 crisis prevented us from fulfilling that intention. Nevertheless, as will be seen, the present sample size still yielded mostly clear-cut results that enabled firm conclusions.

Stimuli and Apparatus

The apparatus and method for creating the stimuli were the same as in Experiment 1.

Procedure

The procedure for the learning phase for the REP and NREP conditions was the same as described in Experiment 1. In the transfer phase, the subjects were instructed to continue to classify the patterns into the same three categories as in the learning phase. In both the REP and NREP conditions, the set of transfer patterns was composed of 15 old distortions (5 per category), 3 prototypes (1 per category), 15 low-level distortions (5 per category), 15 new medium-level distortions (5 per category), and 15 high-level distortions (5 per category). The same procedures for choosing the old distortions in both the REP and NREP conditions were used as in Experiment 1. Each individual pattern was presented once for a total of 63 transfer trials. The order of presentation was randomized for each subject.

Results

We started by conducting preliminary analyses to remove severe outlier subjects. For the learning phase, the performance measure used for identifying outliers was the same as in Experiment 1. For the classification-transfer phase, we measured average accuracy computed across all 63 transfer trials. We again deleted the data of any subject who performed more than 2.5 standard deviations below the mean in each condition on either measure. We deleted 4 subjects from the REP condition (leaving 39 valid subjects) and 2 subjects from the NREP condition (leaving 44 valid subjects). None of our main conclusions changes if all subjects are included in the analyses.

Learning

The results from the learning phase of Experiment 2 are displayed in the bottom panel of Figure 4. The pattern of results is extremely similar to the one in Experiment 1 and provides a close replication of our earlier findings. Most important, learning performance in the REP condition was again far better than in the NREP condition.

We again conducted a 2x15 mixed-model ANOVA using conditions (REP vs. NREP) and blocks as factors. The main effect of learning conditions was significant, F(1, 81) = 18.09 , MSe = 4.356 , p < .001, η2 = .183; as was the main effect of blocks, F(7.14, 578.08) = 56.78, MSe = 1.643, p < .001, η2 = .412. The interaction between the two factors was not significant in this experiment, F(7.14, 578.08) = 1.69 , MSe = .049, p = .107, most likely because the improved performance in the REP condition compared to the NREP condition occurred even more rapidly in Experiment 2 than in Experiment 1.

Transfer– Classification.

To facilitate the presentation, we display the classification-transfer results in two partially overlapping figures: In the middle panel (B1) of Figure 5 we display the probability with which the different types of new transfer patterns (prototype, low distortions, new medium distortions, high distortions) were correctly classified during the transfer phase in the REP and NREP conditions. This figure places focus on the typicality gradient observed for the new transfer patterns. In the bottom panel (B2) of Figure 5, we display the probability with which the old distortions, new medium distortions, and prototypes were correctly classified during the transfer phase in the REP and NREP conditions. This figure places focus on performance comparisons between the old distortions and two of the key new transfer patterns.

As can be seen in the middle panel (B1) of Figure 5, replicating Homa et al., we observed the classic “typicality gradient” in both the REP and NREP conditions, with classification accuracy being highest for the prototypes, followed in order by the low distortions, new medium distortions, and high distortions. We analyzed these data using a 2 x 4 mixed-model ANOVA, with learning condition (REP and NREP) as a between-subject factor and item type (prototype, low, new medium and high distortions) as a within-subject factor. The analysis yielded a main effect of item type, F(2.3, 186.67) = 46.08, MSe = .696, p < .001, η2 = .363, consistent with our observation of the classic typicality gradient. However, there was no main effect of learning condition, F(1,81) = .494, MSe = .030, p = .484. Nor was the interaction between learning condition and item type statistically significant, F(2.3,186.67) = .393, MSe = .006, p = .705. We discuss the null effect of condition more fully in the Modeling section of our article.

As can be seen in the bottom panel (B2) of Figure 5, in the REP condition, the old-medium distortions were classified with higher accuracy than were the new-medium distortions; and were classified with roughly the same accuracy as the prototypes. By contrast, in the NREP condition, the prototypes were classified with the highest accuracy, and there was little if any difference in performance accuracy between the old- and new-medium distortions. To analyze these data, we conducted a 2 x 3 mixed-model ANOVA using as factors learning condition (REP, NREP) and item type (old, new-medium, prototype). The main effect of item type was significant, F(1.62,131.04) = 13.61, MSe = .183, p < .001, reflecting the generally higher performance on the prototypes and old distortions compared to the new medium distortions. There was also a significant condition x item-type interaction, F(1.62,131.04) = 4.72, MSe = .064, p = .016, reflecting the changed accuracy levels of the old distortions compared to the other patterns across the REP and NREP conditions. The main effect of condition was not significant, F(1,81) = 1.82, MSe = .085, p = .181. Subsequent paired-comparison tests showed that the old distortions were classified significantly more accurately than the new medium distortions in the REP condition, t(38) = 5.50, p < .001; although this trend continued to be observed in the NREP condition, the difference was not statistically significant, t(43) = 1.00, p = .646. In addition, the prototypes were classified significantly more accurately than were the old distortions in the NREP condition, t(43) = -2.78, p = .016. That trend was reversed in the REP condition, but the difference in the REP condition was not statistically significant, t(38) = .98, p = .670.

Discussion

Again, consistent with the general qualitative prediction from exemplar models, speed of category learning was significantly faster in the REP condition than in the NREP condition, and the magnitude of the effect was large, averaging 0.136 across the final 8 blocks of learning. The data confirm our pattern of findings from Experiment 1, and are in opposition to Homa et al.’s report of a null effect of the REP/NREP manipulation on speed of category learning in this dot-pattern paradigm.

Consistent with Homa et al.’s findings, the transfer data showed high classification accuracy for all the pattern types in both the REP and the NREP conditions, and also showed the classic “typicality gradient” in both learning conditions, in which patterns with a higher level of distortion from the prototype were classified less accurately. As explained and demonstrated with modeling simulations in our introduction, these classification-transfer findings are consistent with the predictions from exemplar models.

We address in more detail in our subsequent modeling section the extent to which the patterns of classification accuracy for the old distortions compared to the other pattern types can be captured by the exemplar model. In general, however, these results too appear to have a natural account in terms of the model. Because the old distortions receive a very high summed-similarity signal due to their perfect self-match to their repeated representations in memory in the REP condition, the exemplar model naturally predicts that they will be classified more accurately than the new medium-distortions in that condition. By contrast, as explained in the introduction, that self-match contribution to summed similarity is much smaller in the NREP condition (because there is only a single representation of each old test item in memory), and it tends to be swamped by the items’ similarity to all the other training patterns in the NREP condition. Thus, any predicted advantage for the old-medium distortions compared to the new-medium distortions would tend to be quite small in the NREP condition. Finally, predicted accuracy for the prototype tends to be high because it is highly similar to numerous of the old training examples of its category. Indeed, in the NREP condition, the prototype’s summed similarity to the training examples of its category is likely to exceed that of even the individual old distortions themselves. The reason is that although any given old distortion is a perfect match to its single representation in memory, the prototype benefits by tending to have higher similarity to far more other training examples of its category than do the old distortions

Exemplar Model Fit to the Recognition- and Classification-Transfer Data of Experiments 1 and 2

We fitted the same simulation-based exemplar model described in our introduction to the recognition-transfer and classification-transfer data collected in our Experiments 1 and 2. Again, we fitted the model by searching for the values of the free parameters that minimized the sum of squared deviations between the predicted and observed response probabilities for the different item types across both the REP and NREP conditions of both experiments.

To review, the free parameters in the model include 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*, and *c* were held fixed across all experiments and conditions. Separate values of the recognition response-criterion parameter *k* were estimated for the REP and NREP conditions in Experiment 1. (Unlike in our fits to Homa et al.s’ data, we did not need to estimate separate values of *k* for conditions in which foils versus prototypes were tested, because both pattern types were tested within the same transfer test in our Experiment 1.)

Finally, we discovered that noticeably improved fits were achieved when we allowed separate values of the response-scaling parameter γ across the recognition and classification experiments. Although we held this parameter fixed across recognition and classification in our earlier fits to Homa et al.’s data, in hindsight there is no very strong reason to impose this constraint. A process-model interpretation for the response-scaling parameter is that it reflects a criterion for the amount of information that the observer retrieves and accumulates before making a decision (see Nosofsky & Palmeri, 1997, p. 291). There are multiple reasons why this information-accumulation criterion might be expected to differ across the classification and recognition experiments. For example, as explained earlier, classification uses a relative summed-similarity rule whereas recognition uses an absolute summed-similarity rule, so different forms of information are being accumulated and used for making decisions. In addition, in the present paradigm, the observer is choosing among three alternatives in the classification task (category-responses A, B, or C), but is choosing between only two alternatives in the recognition task (old vs. new). In any case, despite allowing separate values of γ across the classification and recognition tasks, the current model is still making use of a relatively small number of free parameters.5

The predictions from the exemplar model are shown as solid dots superimposed on the bars representing observed recognition probabilities in the top panel (A) of Figure 5 and on the bars representing correct classification probabilities in the middle (B1) and bottom (B2) panels of Figure 5, with best-fitting parameters reported in Table 2. Although our main aim involved achieving a reasonable qualitative account of the pattern of results, it can be seen from inspection of the figures that the quantitative fit to the complete set of transfer data is exceptionally good (SSD = .005). All of the major qualitative patterns discussed above for both the classification and recognition transfer data are captured by the model, and usually with high quantitative precision.

The best-fitting parameters (Table 2) showed a similar pattern to the one we reported earlier in fitting Homa et al.’s data (compare to Table 1). Not surprisingly, the between-category distance parameter was estimated to be much greater in magnitude than the within-category distance one. In addition, the recognition-criterion parameter *k* was larger for the REP condition than for the NREP condition. The reason is that subjects tend to set a stricter criterion for the REP condition in response to the generally higher absolute-summed-similarity in the REP condition compared to the NREP condition.

The main difference from our earlier fits is that we allowed separate response-scaling parameters γ across the recognition and classification tasks. For classification, it turns out that γ is estimated at its lower limit of γ=1. The reason has to do with the overall levels of classification accuracy for the novel transfer patterns across the REP and NREP conditions. In general, as γ increases in magnitude beyond 1, the present simulation-version of the exemplar model predicts slightly greater classification accuracy for the novel transfer patterns in the NREP condition than in the REP condition of this dot-pattern paradigm. That data pattern was in fact the trend that Homa et al. (2019) observed in their experiment, although their main effect of condition was not statistically significant. Our classification data went very slightly in the opposite direction (see Figure 5, Panel B1), although again the main effect of condition did not approach statistical significance. It appears that we will require a far larger sample size to pinpoint the true nature of this small predicted effect. The present γ=1 estimate yields predictions of classification accuracy for the novel transfer patterns that are nearly identical across the REP and NREP conditions. Future research is needed to allow us to specify deeper theoretical reasons for the differing γ estimates yielded across the classification and recognition tasks in our experiments.

Finally, as we have already demonstrated in the introduction, a rudimentary learning version of the exemplar model that incorporates “background noise” (Equation 5) would also allow us to capture the general qualitative pattern of results observed in the learning phase of the REP and NREP conditions of our present experiments. However, developing a complete quantitative account of the learning data goes well beyond the scope of the present research. There are undoubtedly an enormous number of complex learning processes that operate in concert, and developing a reasonably adequate model that captures all these learning processes will likely require decades of future research. (For extensive discussion and past attempts at capturing the details of category learning with more complex models with exemplar-based components, see, e.g. Erickson & Kruschke, 1998; Kruschke, 1992; Nosofsky, 1987; Nosofsky, Gluck, et al., 1992; Palmeri, 1997.)

General Discussion

In this article, we addressed a set of challenges to exemplar-based models of classification derived from a version of the classic dot-pattern paradigm tested by Homa et al. (2019). The novel aspect of Homa et al.’s experiments was to conduct versions of the dot-pattern classification-learning paradigm in which training instances never repeated (NREP), and to compare performance in this non-repeating condition to performance in a repeating-training-instances (REP) condition.

The Dissociation Between Classification and Recognition Transfer

Among the purported challenges to exemplar models was that in the NREP condition, classification accuracy was high for all transfer patterns, despite the fact that participants showed zero ability to discriminate between old and new patterns of the same level of distortion during the transfer phase. Here we showed, however, using both conceptual/intuitive lines of argument as well as formal simulation modeling, that this general qualitative pattern of results was in accord with the predictions from exemplar models. In brief, according to the model, classification accuracy was high even in the NREP condition because test patterns from a given category had high similarity to training exemplars from their own category, and low similarity to members of the contrast categories. At the same time, ability to discriminate between old-medium-distortion and new-medium-distortion test items in the recognition task was very poor in the NREP condition, because these items had virtually the same absolute summed similarity to the training examples of the categories (see formal simulation modeling for details). Such “dissociations” between classification and recognition have often been cited as evidence that challenges exemplar models and other closely related models, but careful application of the formal models often reveals that the results are fully compatible with them (e.g., Jamieson, XXX; Love & Gureckis, XXXX; Nosofsky, 1988, 1991, XXXX; Nosofsky & Zaki, 1998). The present example provides another case in point.

It is true that the exemplar model did predict that old patterns should have been recognized with very slightly higher probability than new patterns even in the NREP condition, whereas Homa et al. reported no discrimination ability. Relying on such null effects as a convincing form of evidence, however, requires that researchers provide extensive power-analysis computations or perhaps Bayesian approaches to choosing between null and alternative hypotheses. Such analyses were not part of Homa et al.’s (2019) presentation. As it turned out, in our own replication of Homa et al.’s NREP condition, we found – in contrast to Homa et al. -- that the old-medium distortions were recognized with significantly higher probability than were the new-medium distortions (although tracking down that small predicted difference was not the central purpose of our experiment). Future work aimed at this issue would likely need to conduct experiments with extremely large sample sizes to yield adequate power to detect the very small-size predicted effect.

Finally, beyond accounting for the qualitative dissociation involving classification versus recognition accuracy, we showed that the exemplar model was also able to capture many more fine-grained aspects of the classification and recognition data in both Homa et al.’s (2019) original experiments and our replication experiments, including classification “typicality gradients”, prototype-enhancement effects, and changing patterns of old-new classification and recognition across the transfer phases of the REP versus the NREP conditions.

Classification Learning Across the Repeating and Non-Repeating Conditions

As we acknowledged at the outset of our article, the qualitative result that did most seriously challenge the *a priori* predictions from exemplar models was that, across their three experiments, Homa et al. (2019) observed no difference in the speed of classification learning across the REP and NREP conditions. By contrast, our exemplar-model simulations indicated that learning should have proceeded substantially faster in the REP condition than in the NREP condition. The central motivation for our new experiments was to pursue Home et al.’s null-effect finding involving the learning curves, with the idea of gaining further insights into its nature.

The bottom line, however, is that we observed dramatically different results than did Homa et al. (2019). Across both of our replication experiments, learning proceeded substantially faster in the REP condition than in the NREP condition.

Naturally, we invite independent researchers to conduct the experiment themselves, to perhaps remove experimenter bias from the observed effects. Nevertheless, in our view, our observed positive effect on speed of learning is the intuitively more sensible one. Why wouldn’t learning proceed more rapidly if only a small number of training items needs to be learned and they are repeated over and over again during the course of learning?

We can only speculate about the possible bases for Homa et al. (2019) having observed no effect of the REP vs. NREP manipulation on speed of learning. One possibility may have to do with the detailed nature of their experimental procedures. Homa et al. (2019, p. 398) report that in generating their stimulus materials, “…six different prototypes (A-F) were generated, with about half the subjects receiving one set (A, B, C) and half receiving the other set (D, E, F).” Hence, although presumably randomly generated, the participants in their experiments were exposed to a restricted set of materials. Homa et al. are not explicit about whether the old distortions that participants were exposed to in the REP condition were randomly chosen anew for each and every individual, or whether participants were also exposed to a restricted set of old training distortions. It is possible that, by happenstance, the restricted set of prototypes (and perhaps old training distortions) resulted in difficult-classify patterns in the REP condition, and easier-to-classify patterns in the NREP condition. By contrast, in our procedure, the materials were generated randomly anew for each and every participant in each and every condition. In our view, our procedure is the preferable one if one is seeking generality and robustness in identifying the phenomena of interest. Conceivably, the restricted set of materials used by Homa et al. could also have influenced the expected magnitude of the old-new recognition effect in the transfer phase of their NREP condition as well.

Modeling Comparisons

As noted in the introduction section of our article, Homa et al. (2019) also reported model-based analyses of their classification and recognition data. We have already pointed out potential limitations of their approach to estimating and inserting “average-similarity” parameters into the prediction equations for classification and recognition. More generally, however, we need to point out that Homa et al. never actually presented the overall pattern of predictions from their version of the exemplar model; instead, they presented the exemplar-modeling results in terms of plots reporting deviations between predicted and observed classification and recognition probabilities for the individual item types (see Homa et al., 2019, Figure 14). Thus, the overall qualitative patterns of predictions yielded by the model were not clearly illustrated to readers. Moreover, the deviations in their Figure 14 are plotted on a magnified scale ranging from a-.04 to .02 (or from -.02 to .04), so even small deviations can appear to be large in this plot.

In any case, although the goals of our modeling in the present article were to capture only the main qualitative pattern of results in the data (for reasons we explained in the Model Development section), the quantitative fits yielded by the exemplar model were better than those reported by Homa et al. (compare deviations in our model-prediction graphs to Homa et al.’s Figure 14). Beyond the use of different formal instantiations of the exemplar model, a possible reason is that Homa et al. constrained their models to simultaneously predict the classification learning data and the classification and recognition transfer data. In doing so, they were forcing the model to try to fit overlapping learning curves across the REP and NREP conditions, a pattern of results that goes completely outside the character of what the model actually predicts. Our data suggest that forcing the model to fit the overlapping learning curves introduced an inappropriate constraint; therefore, in our view, the exemplar-model fits that Homa et al. reported for their entire data set must be interpreted with caution.

The Bigger Picture

In concluding, we need to make clear that in this article we have focused on challenges to exemplar models derived from the highly influential dot-pattern classification-learning paradigm and the recent version of the paradigm advanced by Homa et al. (2019). We believe we have addressed these most recent challenges in an effective manner in this article. However, we are *not* claiming that exemplar models of classification and recognition have not received serious challenges from other sources. Although exemplar models appear to have a great deal of generality and are often serious contenders, researchers have of course identified numerous limitations associated with such models in other paradigms. A full survey goes way beyond the scope of this article, but a brief listing of some interesting examples include base-rate neglect effects (Gluck & Bower, 1988); rule-extrapolation effects (Erickson & Krushke, 2002); category-contrast sequence effects (Stewart, Brown, & Chater, 2002); response-time evidence for sequential logical-rule application (Little, Nosofsky, & Denton, 2011); category-structure-learning effects (Conaway & Kurtz, 2017); and extreme prototype-enhancement effects in recognition-memory experiments involving stimuli with high sensory noise (Dube, 2019). Scientific progress often proceeds step-by-step, and it is important to identify which specific challenges provide convincing evidence of limitations of proposed models, and which do not. We have taken only one such important step in the present research.

References

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.

Conaway, N., & Kurtz, K. J. (2017). Similar to the category, but not the exemplars: A study of generalization. *Psychonomic bulletin & review*, *24*(4), 1312-1323.

**Dubé, C.** (2019). Central tendency representation and exemplar matching in visual short-term memory. Special Issue to Commemorate the 50th Anniversary of Atkinson and Shiffrin, Memory and Cognition, 47, 598-602.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107.

Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, *9*(1), 160-168.

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.

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

Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 1.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 324.

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.

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.

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.

Stewart, N., Brown, G. D., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 3.

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.

3. The Greenhouse-Geisser correction was applied for violation of the sphericity assumption.

4. In this article, *p* values of multiple t tests conducted on the same data set were adjusted using Bonferroni correction. If any *p* value is less than .05 before the correction but greater than .05 after the correction, we refer to the effect as “marginally significant”.

5. We should emphasize that the fits are still very good if we do constrain γ to be fixed across classification and recognition: total SSD=.018 with 99.6% of the variance accounted for across all items types in the REP and NREP conditions in both the recognition-transfer and classification-transfer experiments.

Table 1. Best-Fitting Parameter Values from the Exemplar Model Fits to the Recognition- and Classification-Transfer Data of Homa et al.’s (2019) experiments.

|  |  |
| --- | --- |
| Parameter | Value |
| *between* | 3.181 |
| *within* | 0.206 |
| *c* | 0.696 |
| *γ* | 1.620 |
| *k* (REP, Exp. 2) | 21.119 |
| *k* (NREP, Exp. 2) | 10.474 |
| *k* (REP, Exp. 3) | 26.956 |
| *k* (NREP, Exp. 3) | 11.979 |

Note. *between* = between-category dissimilarity; *within* = within-category dissimilarity; *c* = sensitivity; γ = response-scaling; *k* = response criterion. Separate values of *k* are estimated for each of the REP and NREP conditions in Experiments 2 and 3.

Table 2. Best-Fitting Parameter Values From the Exemplar Model Fits to the Recognition- and Classification-Transfer Data of Experiments 1 and 2.

|  |  |
| --- | --- |
| Parameter | Value |
| *between* | 4.000 |
| *within* | 0.301 |
| *c* | 0.491 |
| *γ* (Recognition) | 1.922 |
| *γ* (Classification) | 1.000 |
| *k* (REP, Exp. 1) | 51.547 |
| *k* (NREP, Exp. 1) | 13.526 |

Note. *between* = between-category dissimilarity; *within* = within-category dissimilarity; *c* = sensitivity; γ = response-scaling; *k* = recognition response criterion. Separate values of γ are estimated for the recognition and classification tasks. Separate values of *k* are estimated for REP and NREP conditions of Experiment 1.

Figure Captions

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.

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 (black dots) are from the simulation-based version of the exemplar model described in the text.

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.

4. Mean proportion of correct classifications as a function of the number of blocks for the REP and NREP conditions. Top panel: Experiment 1, bottom panel: Experiment 2.

5. Observed (colored bars) and predicted (black dots) proportions of recognition and classification responses for the different item types in the REP and NREP conditions across Experiments 1 and 2. Panel A: old response probabilities for the four different types of transfer patterns (old, new medium, prototype, foil) in the REP and NREP conditions, Experiment 1. Panel B1: Correct classification probabilities for the novel transfer patterns (prototype, low, medium, high) in the REP and NREP conditions, Experiment 2. Panel B2: Correct classification probabilities for the old distortions, new medium distortions, and prototypes in the REP and NREP conditions, Experiment 2. Note: In Panel B1 the different colors denote learning conditions, whereas in Panel B2 the different colors denote item types. Predictions (solid dots on each bar) are from the simulation-based exemplar model described in the text. Error bars denote standard error of the mean.

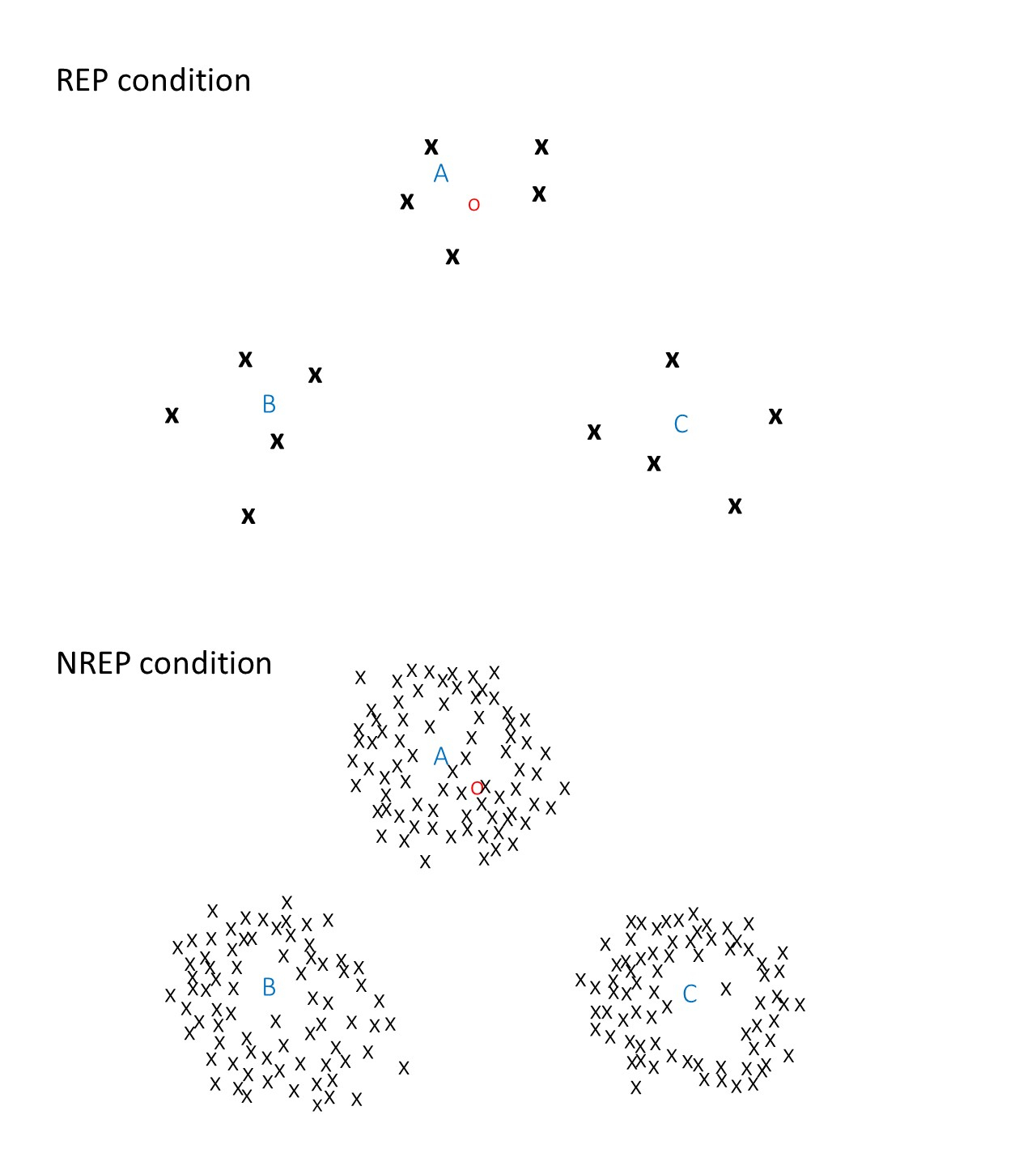


Figure 1

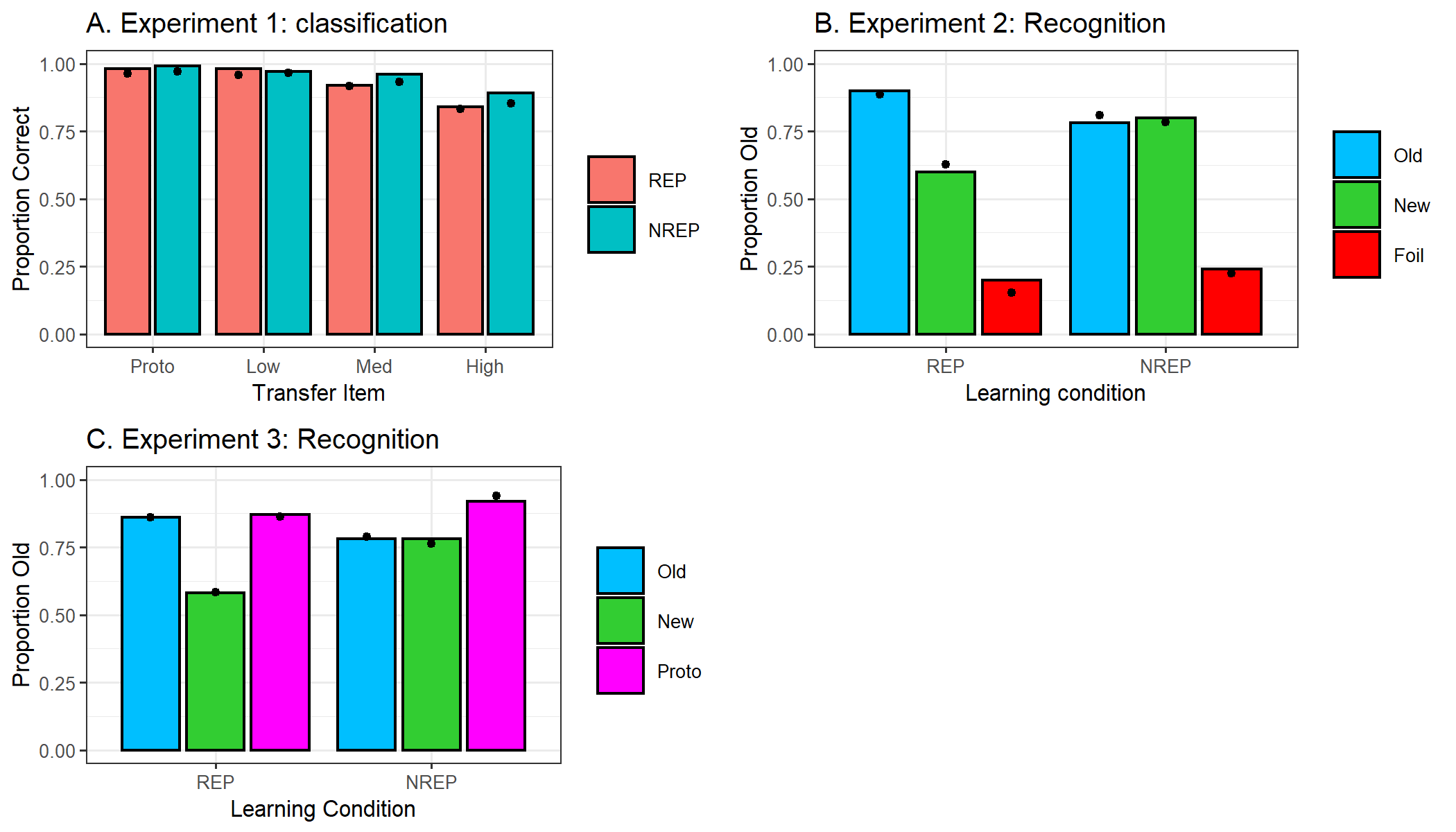


Figure 2.

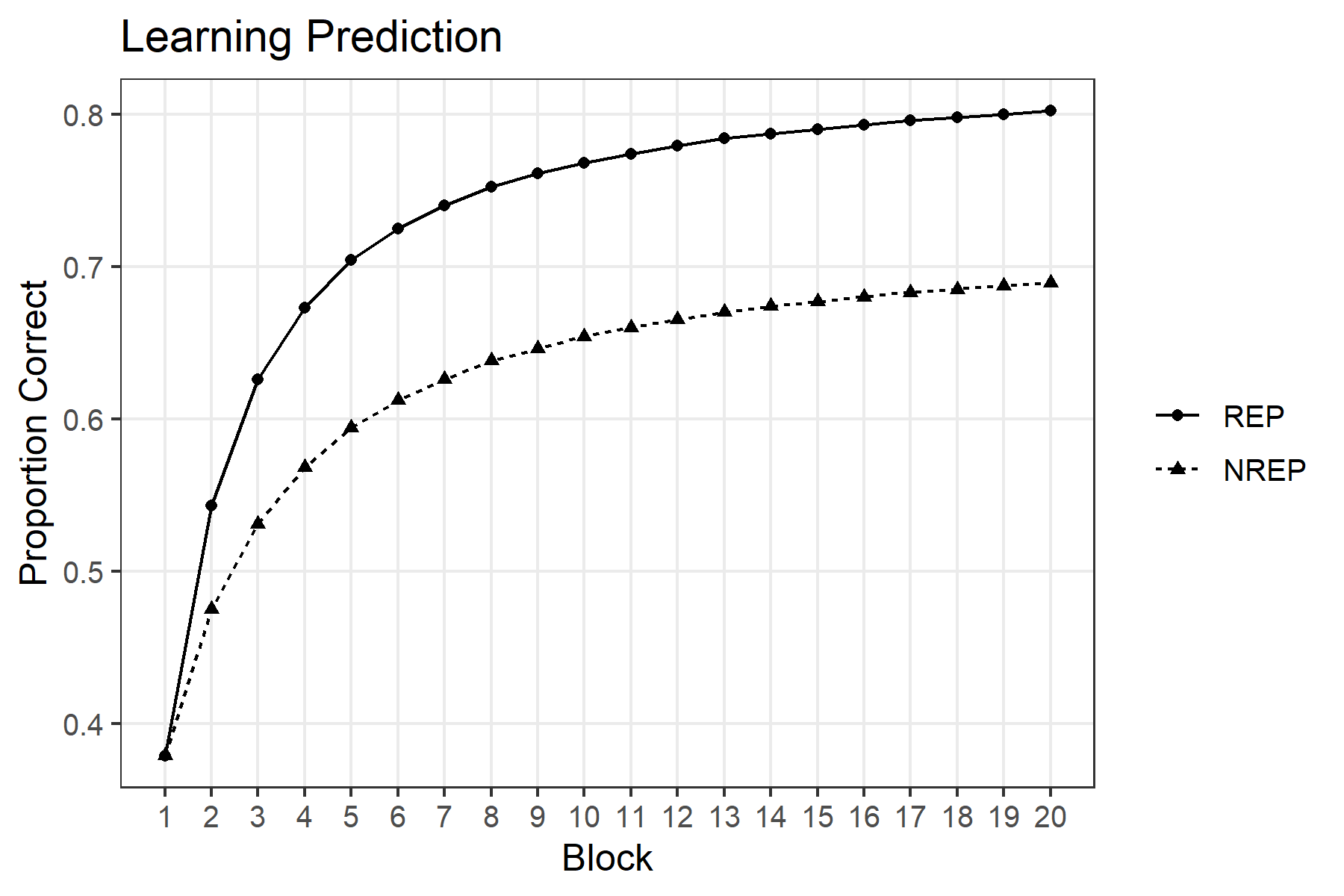
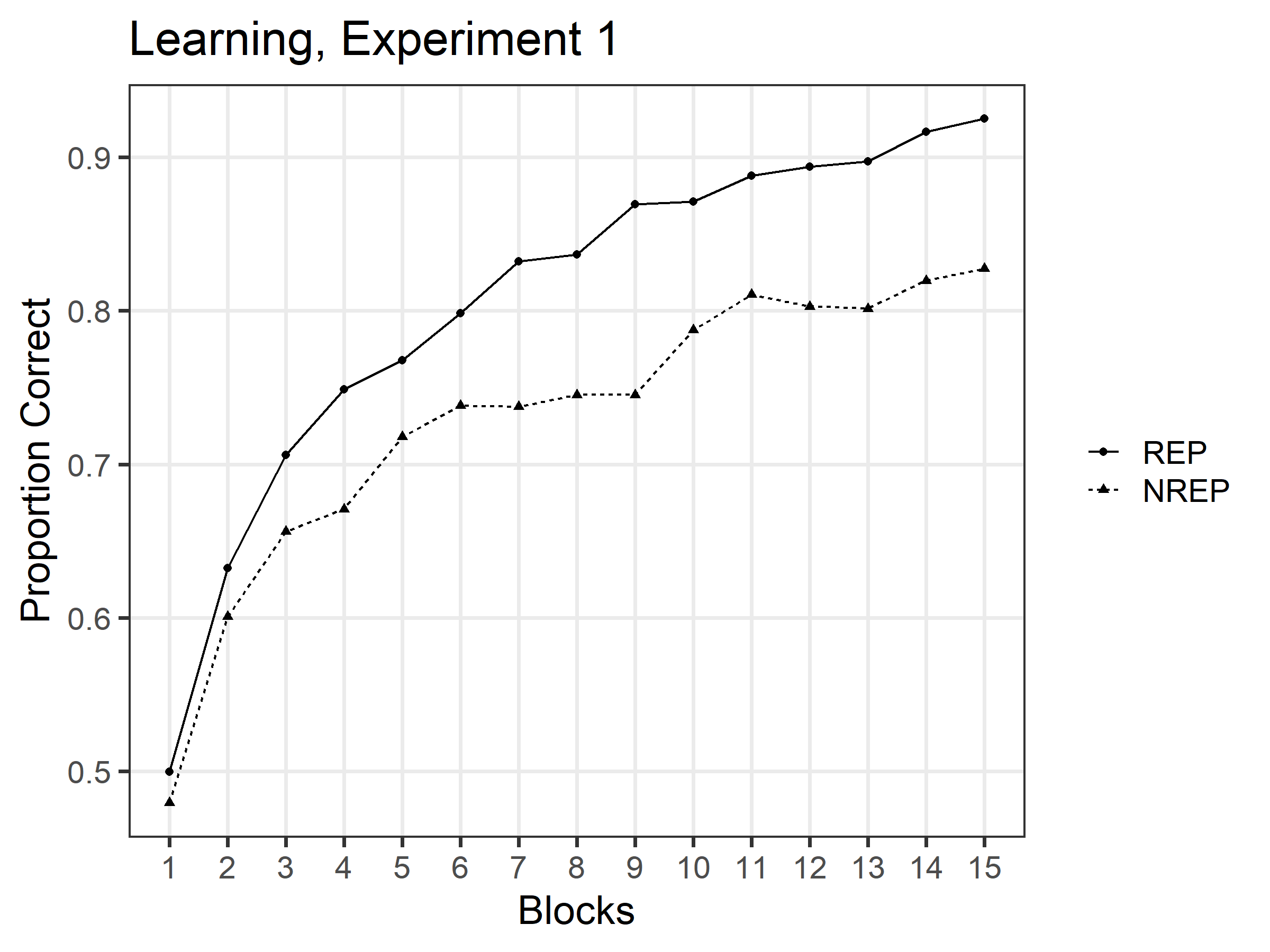


Figure 3.



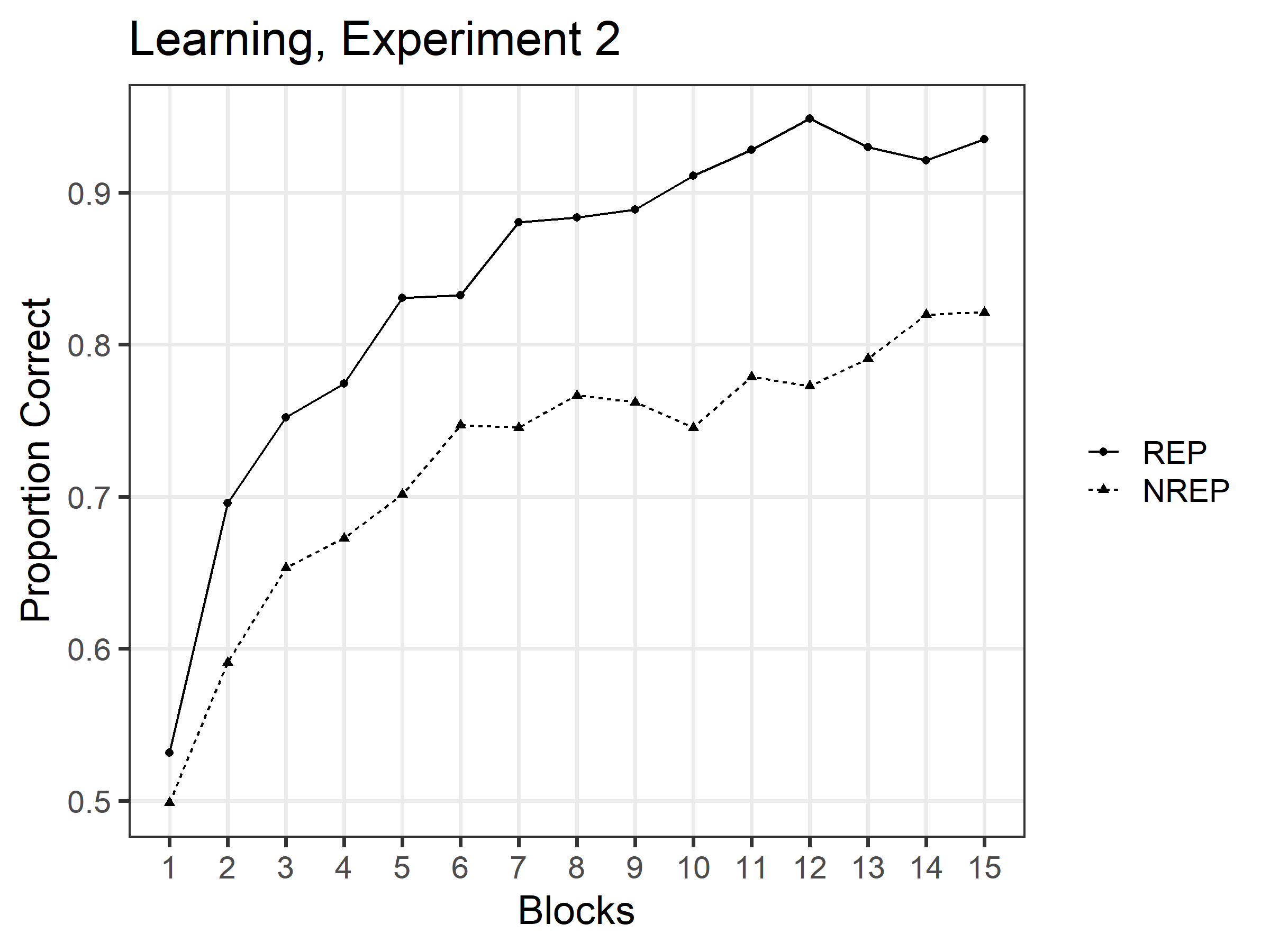


Figure 4.

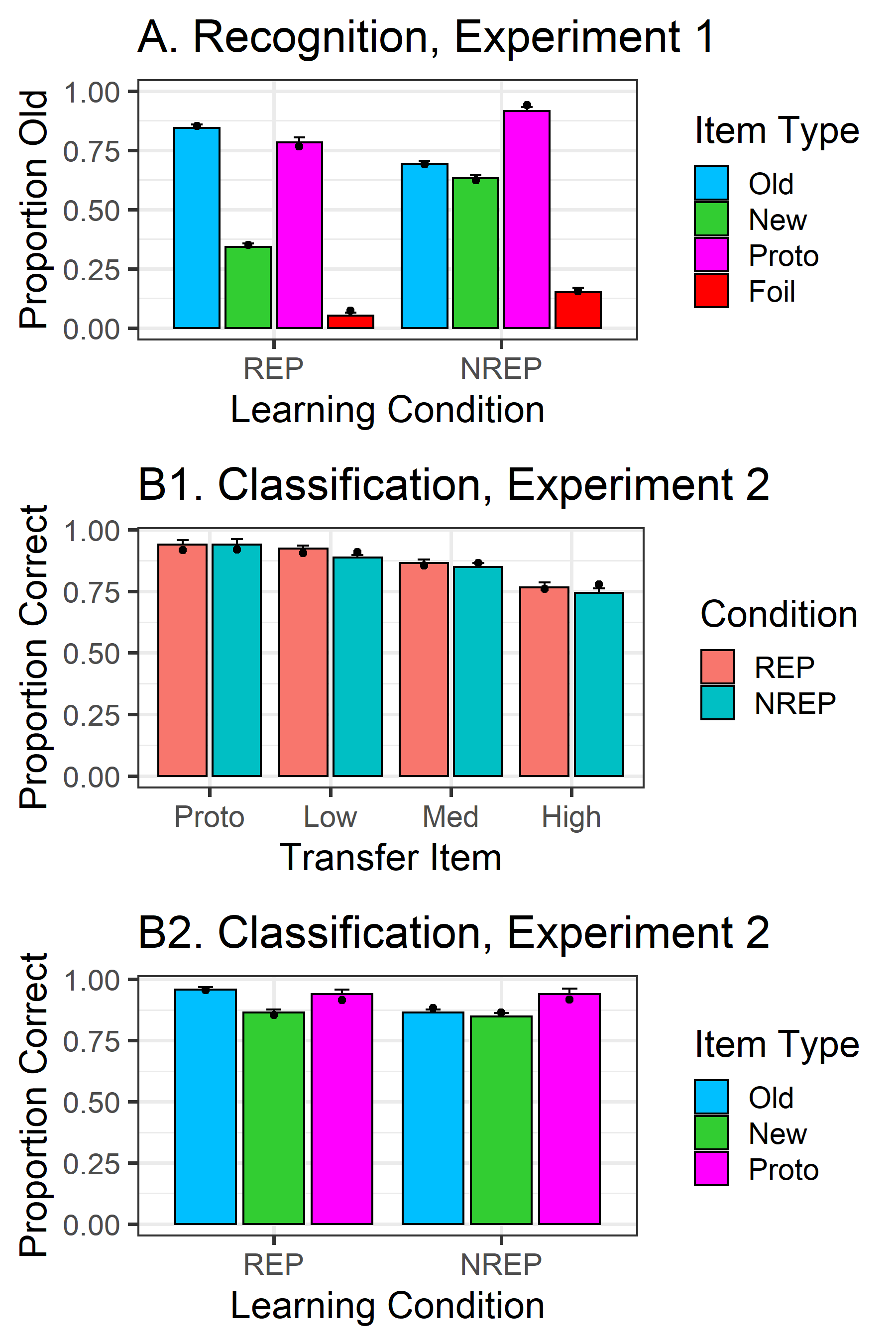


Figure 5.