Experiment 1

The purpose of experiment 1 was to replicate the learning and recognition-transfer phases as in Homa et al.’s experiments 2 and 3. The structure of the learning phases in the REP and NREP conditions was the same as Homa et al.’s experiments. To investigate whether the speed of classification learning is different or not across the two learning conditions, we measured the classification accuracy for each learning block in both conditions. The transfer phase involved recognition tests on old-medium distortions, new-medium distortions, prototypes, and foils. The recognition performance for the prototypes and foils was assessed in a single experiment so as to increase the statistical power. We expected the pattern of recognition data to be consistent with Homa et al.’s results as well as the qualitative predictions of the exemplar model, mainly in that the old-new discrimination would be excellent in the REP condition but very poor in the NREP condition.

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. Each pattern consisted of 9 dots positioned in the central 30 × 30 area of a 50 x 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 white patterns were displayed at the center of a grey computer screen. {We should say things such as that the patterns were displayed centered on the computer screen, and provide their rough size and visual angle.}

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 training 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 pattern 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 Figure 1. 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 analysis revealed a significant 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.

\*Greenhouse-Geisser correction applied for violation of the sphericity assumption.

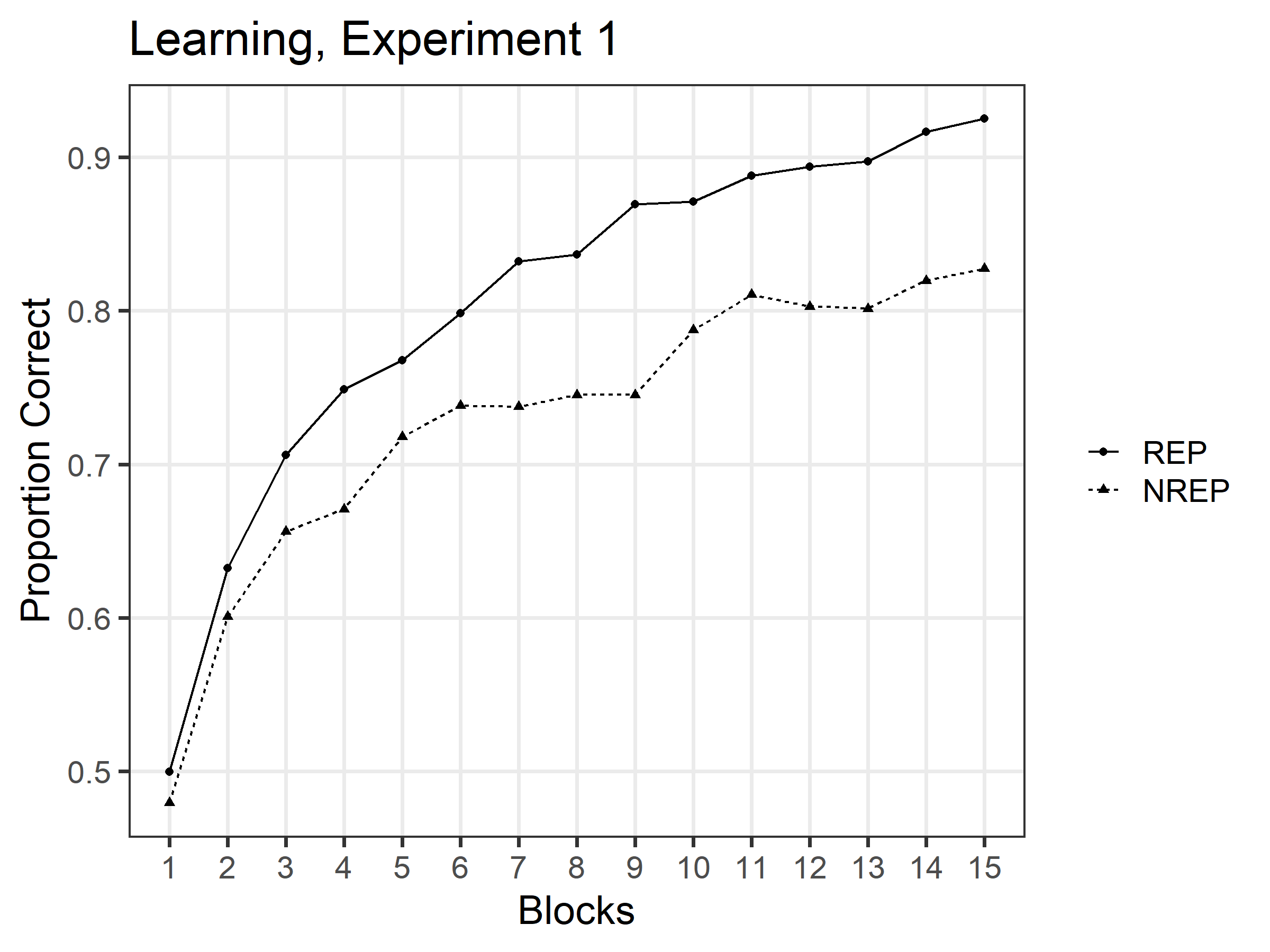


Figure 1 Mean proportion of correct classifications as a function of the number of blocks for the REP and NREP conditions, Experiment 1.

Transfer-Recognition.

The probability with which each type of transfer pattern was judged as old in the REP and NREP conditions is shown in Figure 2. 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\*. 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.

\*In this paper, p values of multiple t tests conducted on the same data set were adjusted for 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”.

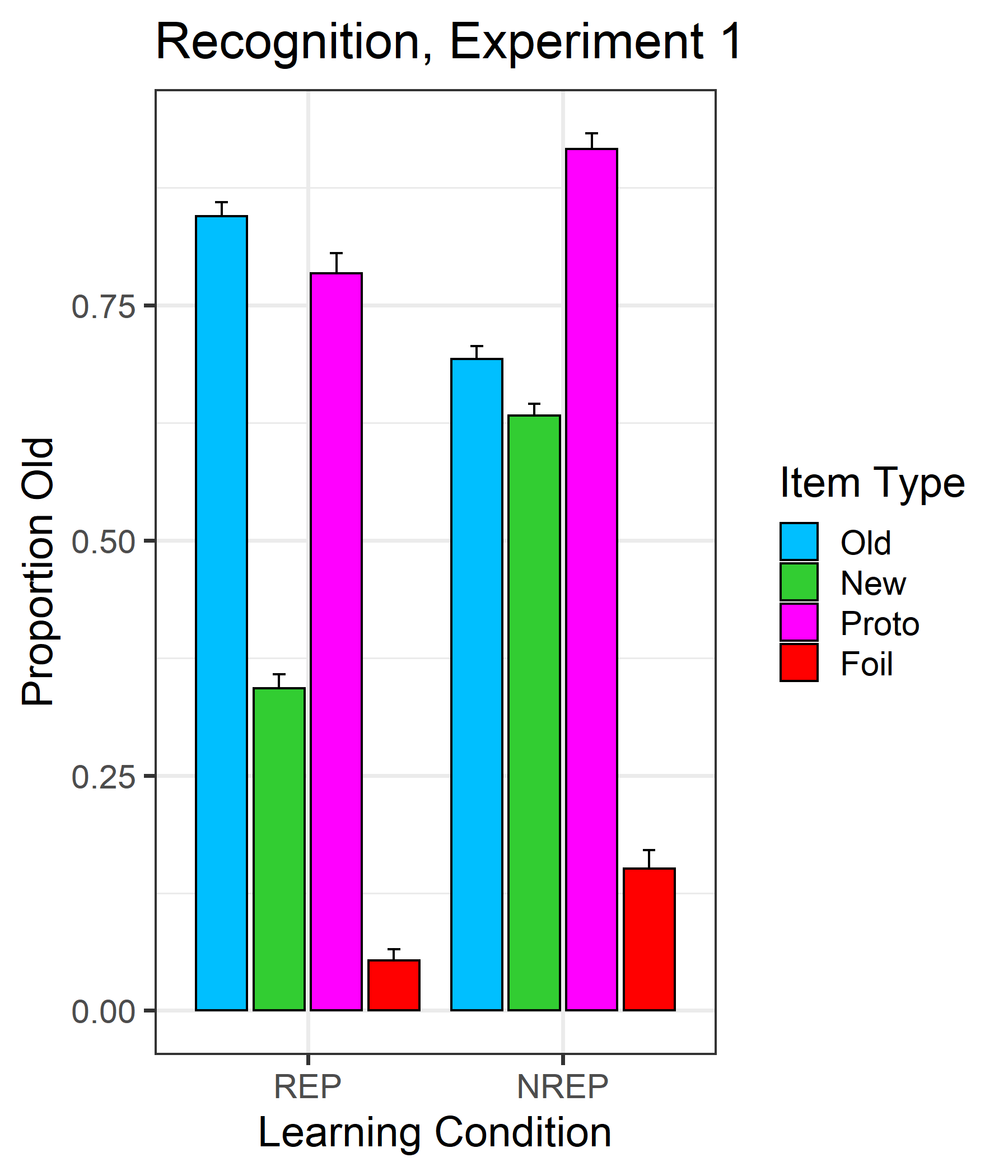


Figure 2 Mean proportion of old responses (with standard error bars) for the four different types of transfer patterns (old, new medium, prototype, foil) for REP and NREP conditions in Experiment 1.

Discussion

Contrary to Homa et al.’s finding, the learning data showed that the speed of learning was faster in the REP condition than in the NREP condition, implying that the category learning was facilitated by having some patterns repeated in each learning block. Consistent with Homa et al.’s findings, the transfer data showed that subjects easily discriminated old and new patterns in the REP condition but had difficulty discriminating old and new patterns in the NREP condition, implying that memory strength for training patterns was greatly enhanced by repeated exposure to a fewer number of patterns. The trend in the false alarm rates of the prototypes and foils was also perfectly replicated: the prototypes were almost as likely to be judged as old as old-medium distortions in the REP condition, and was even more likely to be mistaken as old patterns than actual old-medium distortions in the NREP condition; in addition, the false alarm rate of the foils was quite low regardless of the learning conditions.

Experiment 2

Experiment 2 was intended to replicate the learning phase and transfer-classification phase, as in Homa et al.’s experiment 1. The structure of the learning phases in the REP and NREP conditions was similar to Homa et al.’s experiment 1, except that there were 15 learning blocks in our experiment. The learning phases in the REP and NREP conditions were replicated to confirm our finding in experiment 1 that the classification learning rate was faster in the REP than in the NREP condition. The transfer phases involved classification tests on prototypes, old-medium, low, new-medium, and high distortions. The old-medium distortions were tested in addition to the four item types used in the Homa et al.’s experiment 1 in order to constrain the simulation process of the exemplar model. We expected the pattern of classification data to be consistent with Homa et al.’s results as well as the qualitative predictions of the exemplar model, mainly in that the classification performance of the novel test patterns would be excellent in both the REP and NREP conditions.

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.

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. The measure of learning-phase performance was the same as in Experiment 1. For the classification-transfer phase, we measured average accuracy computed across all 69 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 Figure 3. The pattern of results is extremely similar to the one in Experiment 1 and provides a close replication of the earlier findings. Most important, learning performance in the REP condition was again clearly 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.

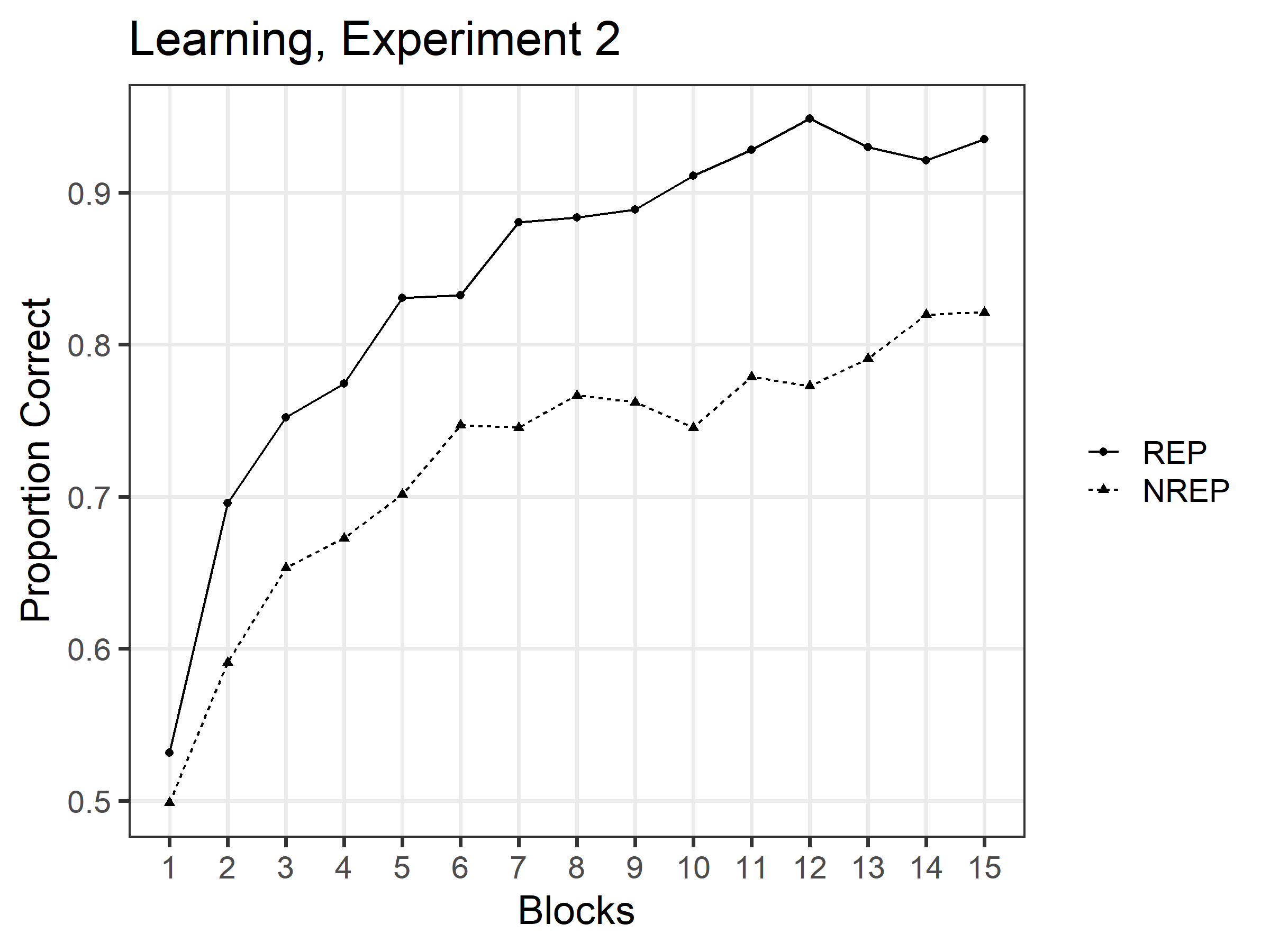


Figure 3 Mean proportion of correct classifications as a function of blocks for the REP and NREP conditions in Experiment 2

Transfer– Classification.

To facilitate the presentation, we display the classification-transfer results in two partially overlapping figures: In Figure 4 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 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 Figure 4a, 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], confirming 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].

As can be seen in Figure 4b, 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 differences in the REP condition was not statistically significant, t(38) = .98, p = .670.

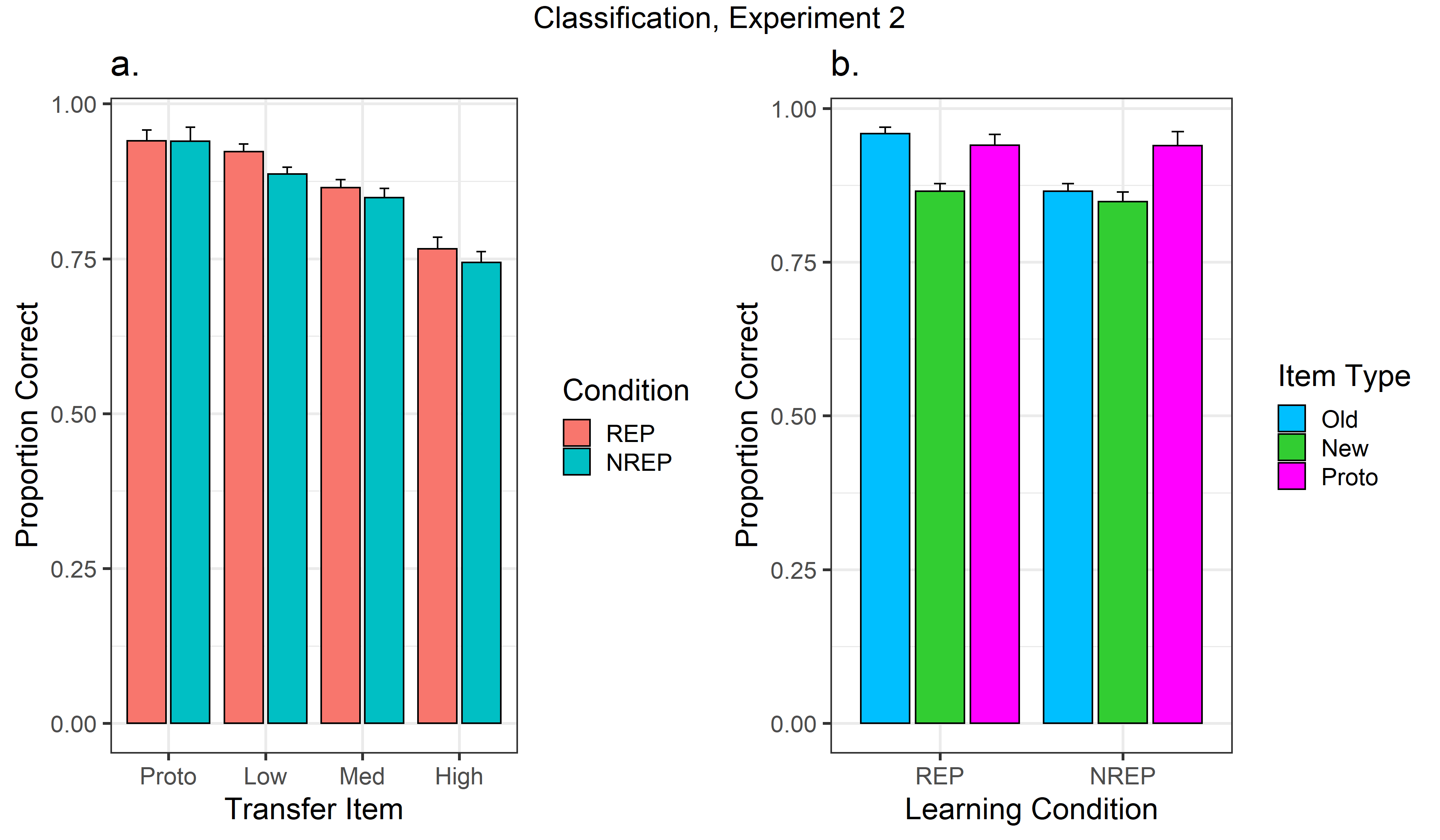


Figure 4 Mean proportion of correct classifications (with standard error bars) to five different types of transfer patterns (old medium distortion, prototype, low distortion, new medium distortion, high distortion) for REP and NREP conditions, Experiment 2. In panel a, two different colors denote the REP and NREP conditions. In panel b, three different colors denote the old, new-medium distortions and prototypes.

Discussion

Again, the learning data suggested that the category learning was facilitated by having some patterns repeated in each learning block. 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. It is worth noting that the transfer patterns were classified very accurately despite the fact that no single training pattern was repeated during the learning phase and that subjects can hardly discriminate between the old- and new-medium distortions during the recognition transfer test. We also confirmed the “typicality gradient” in both learning conditions as in Homa et al.’s experiment 1, in which novel patterns with a higher level of distortions from the prototype were classified less accurately. Moreover, we observed that the endorsement rates of the prototypes tended to be very high in both conditions, even exceeding that of the old distortions in the NREP condition, probably due to the existence of numerous highly similar training patterns to the prototype in the NREP condition. Lastly, as can be inferred from the trend in the recognition performance, the classification accuracy of the old-medium distortions was higher than that of the new-medium distortions in the REP condition, when subjects can easily discriminate between the two pattern types. In contrast, the classification accuracy of the old-medium distortions was virtually identical to that of the new-medium distortions in the NREP condition, when subjects can hardly discriminate between the two pattern types. Therefore, the stronger memory traces for old-medium distortions in the REP condition enhance the classification rate of these old distortions.

Exemplar model fit to the recognition and classification data

We fitted the same simulation-based exemplar model that was used to fit the Homa et al.’s classification and transfer data to these data in our experiments by searching for the free parameters in the model that minimized the sum-of-squared deviations between the predicted and observed probabilities. As a review, free parameters contained 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 response-criterion parameter k were estimated for each of the REP and NREP conditions in experiment 2. Separate values of the response-scaling parameter γ were estimated for each of the classification and recognition experiments but held fixed across both conditions in each experiment, to account for the difference in the decision rules involved in the classification and recognition judgments. γ was constrained to be greater or equal to 1 such that subjects were always assumed to classify a test item to its target category at least with a probability dictated by the relative-summed-similarity.

The predictions from the exemplar model are shown as solid dots in Figure 5, with best-fitting parameters reported in Table x. It turns out that the quantitative fit to the data is exceptionally well (SSD = .005). All of the major qualitative patterns discussed above for both the classification and recognition data are captured by the model, and usually with high quantitative precision.

             As can be seen from panel A of figure 5, the exemplar model predicts a huge difference between the old-recognition probabilities for the old- versus new-medium distortions in the REP condition, yet little difference between the two pattern types in the NREP condition. It was also predicted that the false alarm rates for the prototypes were relatively high in both conditions, with the rate exceeding the old-recognition probabilities of old-medium distortions in the NREP condition; and that the false alarm rates for the foils in both conditions were extremely low. All these results align with the trends in our recognition data as discussed earlier.

             As can be seen from the panel B1, the overall classification accuracy was high for each of the pattern types in both conditions. The “typicality gradient” observed in our classification data was also predicted for patterns with varying levels of distortion from their prototypes in each of the two conditions. Surprisingly, for each of the four pattern types, the exemplar model predicted slightly higher classification accuracy in the NREP condition than in the REP condition. However, the classification data of our experiment 2 showed the opposite pattern. Nevertheless, the predicted differences in the classification accuracies between the REP and NREP conditions were so small that measurements from a large subject pool were required to detect it. In fact, none of the between-condition differences observed in the classification data was significant. In addition, the subtle interaction effects between three item types (old, new-medium and prototype) and two learning conditions (REP and NREP) demonstrated in the panel B2, as discussed earlier, are also predicted by the exemplar model with high quantitative precision.

Inspection of the best-fitting parameters (Table x) revealed that the between-category distance was estimated to be much greater than the within-category distances. In addition, the recognition-criterion parameter k was larger for the REP condition than for the NREP condition in experiment 1. 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. Moreover, the response-scaling parameters γ were indeed estimated to be different between the two experiments, suggesting different decision strategies utilized in different experimental paradigms. We speculated two reasons why subjects adopted different decision strategies: First, there are only two options, old or new, in the recognition experiment, whereas there were three options for categorizing a test item into one of the three categories experienced. Second, absolute-summed-similarity was evaluated in a recognition judgment, but relative-summed-similarity was evaluated in a classification judgment. The specific values of the response-scaling parameters revealed that, on average, subjects responded by probability-matching in the classification experiment and responded more deterministically towards the option with stronger evidence in the recognition experiment.

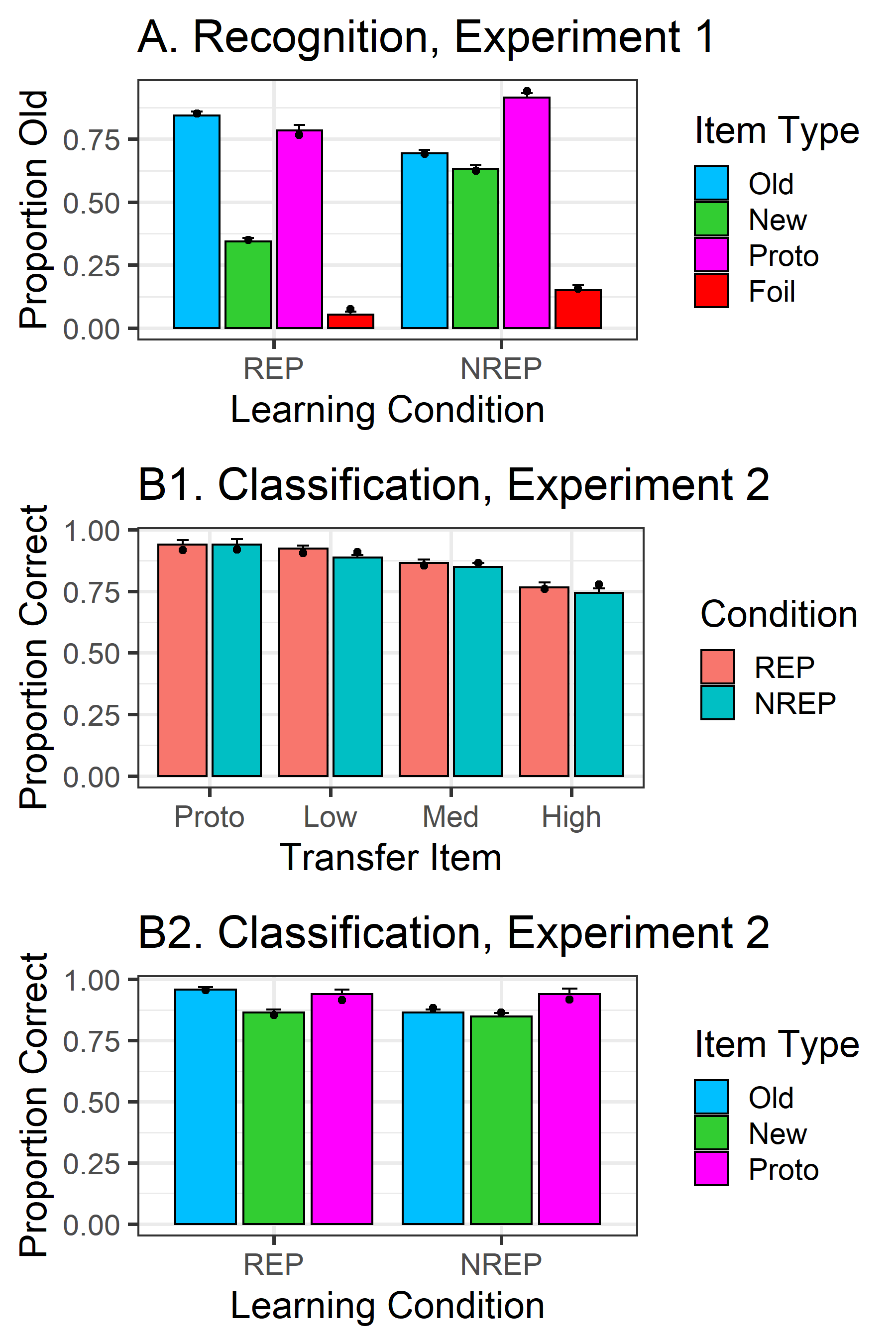


Figure 5 observed and predicted probabilities of old responses in experiment 1 and of correct classifications in experiments 2 for each item types, shown for the REP and NREP conditions separately. The colored bars represent observed data and the solid dots on each bar represent data predicted by exemplar model.