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. Each pattern consisted of 9 dots 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. {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 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 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.

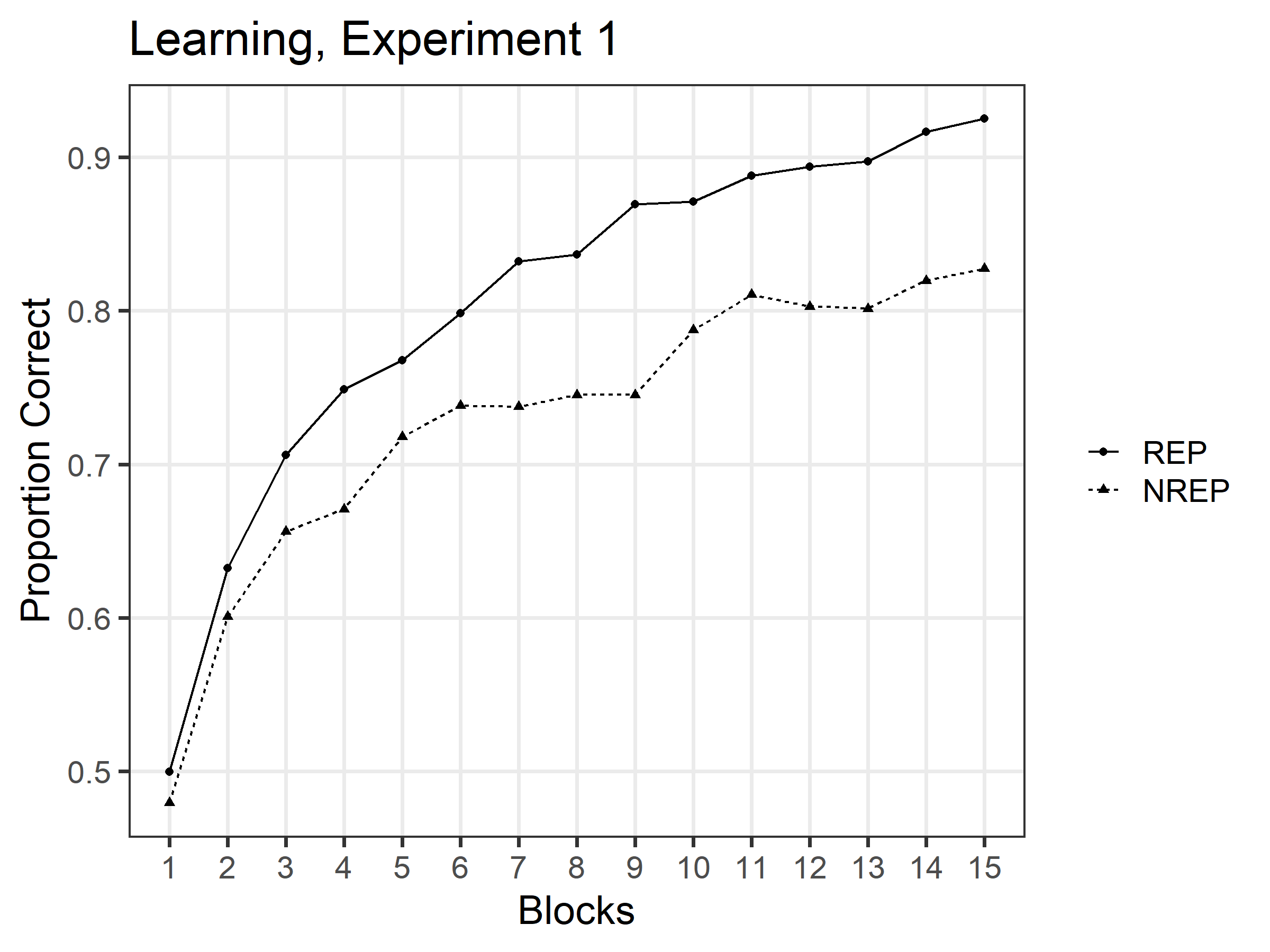


Figure 1 Mean proportion of correct classifications as a function of the number of blocks for the REP and NREP conditions, Experiment 1.

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”.

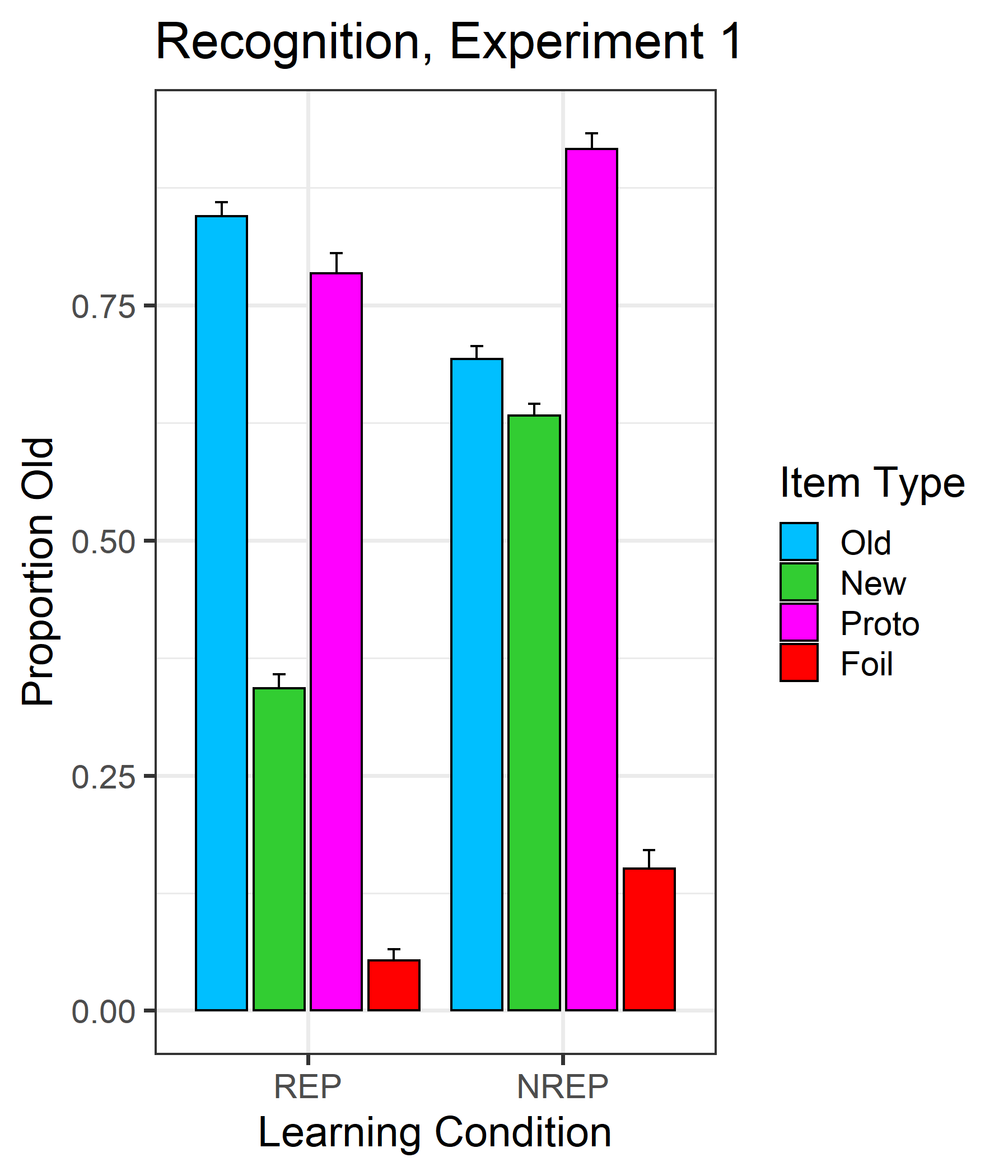


Figure 2 Mean proportion of old responses (with standard error bars) for the four different types of transfer patterns (old, new medium, prototype, foil) for REP and NREP conditions in Experiment 1.

Discussion

Contrary to Homa et al.’s 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 XXX 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 Figure 3. 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.

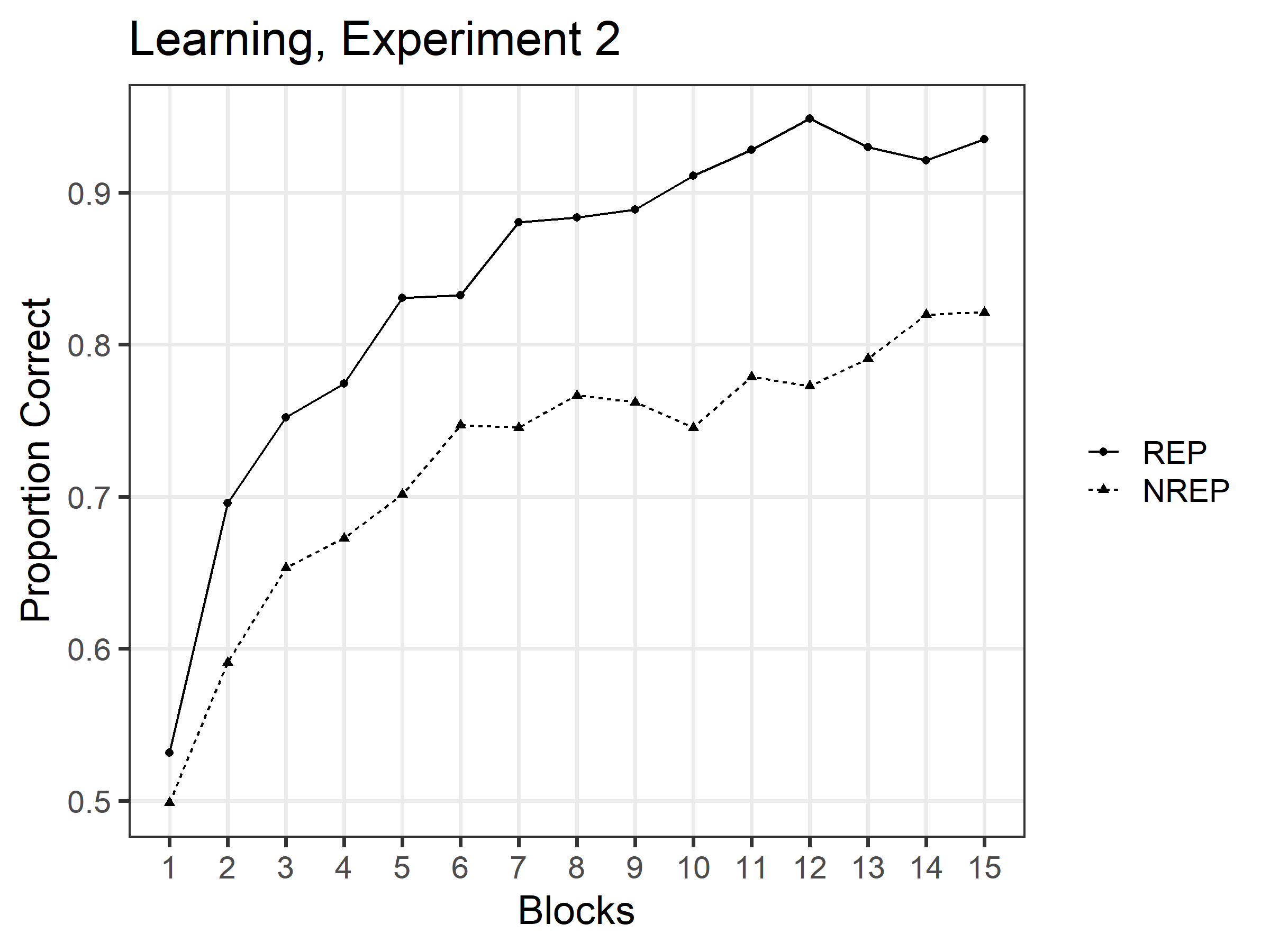


Figure 3 Mean proportion of correct classifications as a function of blocks for the REP and NREP conditions in Experiment 2

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, 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 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 difference in the REP condition was not statistically significant, t(38) = .98, p = .670.

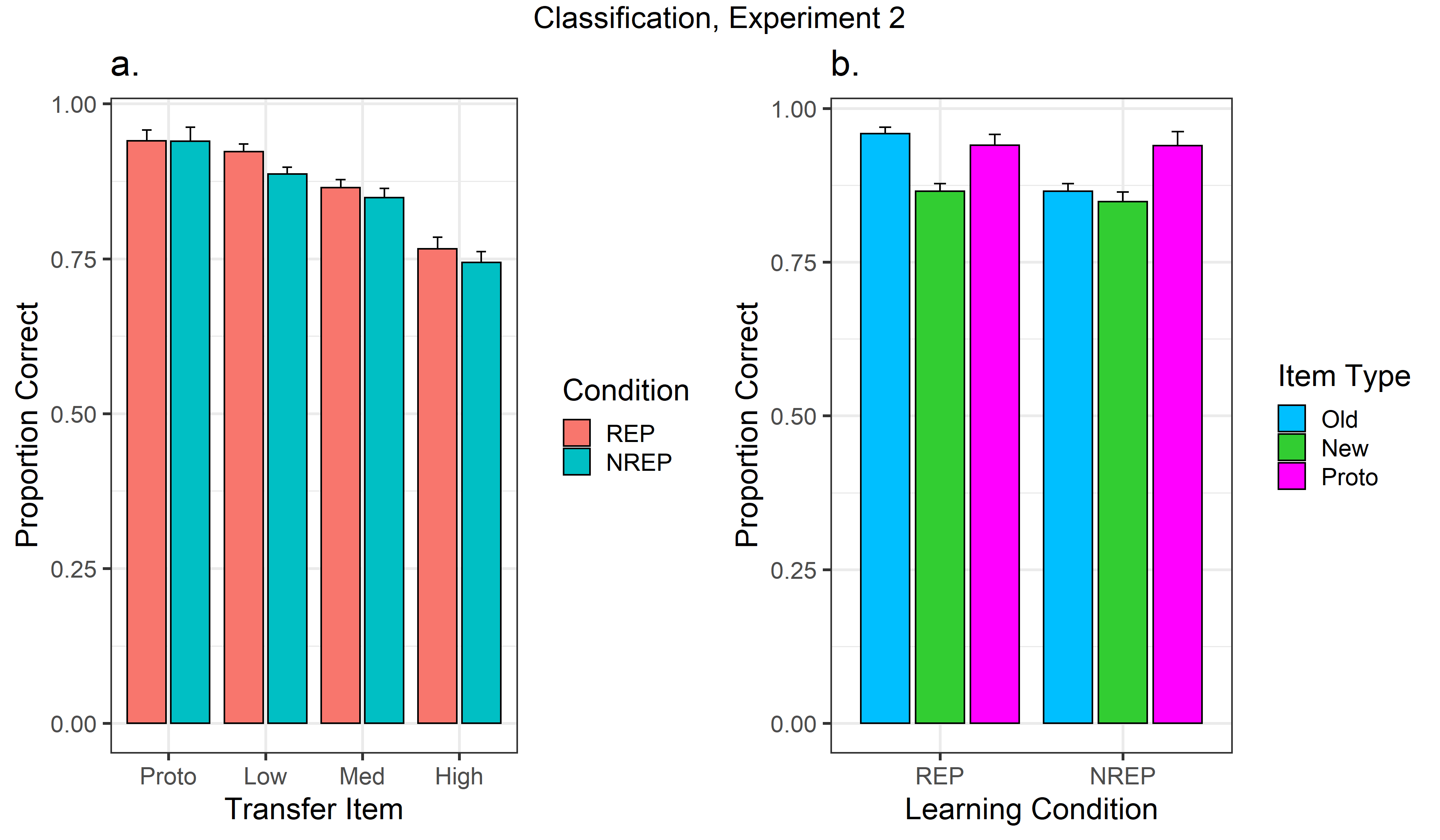


Figure 4 Mean proportion of correct classifications (with standard error bars) to five different types of transfer patterns (old medium distortion, prototype, low distortion, new medium distortion, high distortion) for REP and NREP conditions, Experiment 2. In panel a, two different colors denote the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

Discussion

Again, 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 XXX 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 classification-transfer and recognition-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, pp. XXXX). 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. {Footnote: We should emphasize that the fits are still reasonably good if we do constrain γ to be fixed across classification and recognition: total SSD=.018 with XX.X% of the variance accounted for across all items types in the REP and NREP conditions in both the recognition-transfer and classification-transfer experiments.}

The predictions from the exemplar model are shown as solid dots in Figure 5, with best-fitting parameters reported in Table x1. 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 x1) showed a similar pattern to the one we reported earlier in fitting Homa et al.’s data (compare to Table x2). 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 slightly in the opposite direction (see Figure z), 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.

As we have already demonstrated in the introduction, a rudimentary learning version of the exemplar model that incorporates “background noise” 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 beyond the scope of the present research, as there are undoubtedly an enormous number of complex learning processes that operate in concert (for 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).

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

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

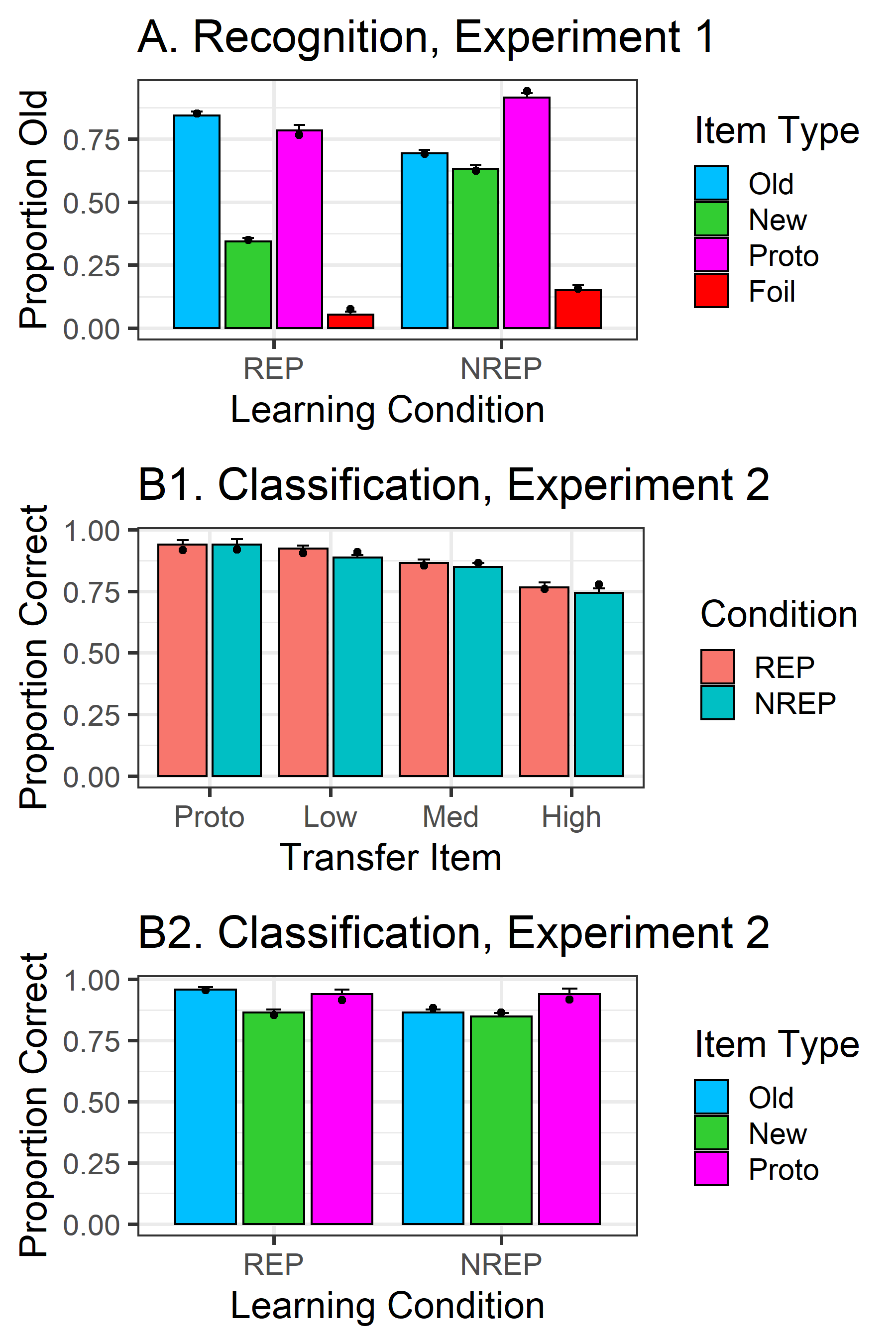


Figure 5 observed and predicted probabilities of old responses in experiment 1 and of correct classifications in experiments 2 for each item types, shown for the REP and NREP conditions separately. The colored bars represent observed data and the solid dots on each bar represent data predicted by exemplar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