Training Methods in Categorization: A Comparison of Classification and Observation on Rule Adoption and Rule Consistency

Yu-Wei Chang (ychang01@syr.edu)
Michael L. Kalish (mlkalish@syr.edu)
Daniel Corral (dcorral@syr.edu)
Department of Psychology, Syracuse University,
430 Huntington Hall, Syracuse, NY 13244 USA

Abstract

This study compares classification and observation training in categorization tasks involving multiple rules, including an optimal XOR rule and suboptimal uni-dimensional rules. Participants (N = 192) were assigned to either condition, with classification involving active categorization and feedback, while observation involved studying the pair of category label and item together. Results showed that classification participants outperformed observation participants in accuracy and exhibited greater consistency in strategy use. Bayesian modeling revealed no significant difference in rule adoption between conditions, but classification led to fewer strategy switches and lower error rates. These findings suggest that classification training enhances performance by fostering stronger commitment to adopted strategies. The study highlights the importance of strategy commitment in categorization and questions the reliance on overall accuracy alone as a performance metric.

Keywords: Categorization; Classification Training; Observational Training; Strategy Identification.

Introduction

Previous studies have demonstrated that individuals can learn categories through different methods of presenting the categorical information. In this study, we compare the two most common training methods: classification and observation.

Classification Versus Observation

In classification training, participants are presented with an item and asked to classify it into one of several categories. After making their classification judgment, participants receive *corrective feedback*—participants are shown the correct answer and whether their response was correct—on their answer and are encouraged to learn from their mistakes. In contrast, during observational learning, participants are shown the category label along with the item and are asked to study and learn the categorization.

A great body of research has examined the difference between classification and observation training; however, the results remain inconsistent across studies. Ashby et al. (2002), Edmunds et al. (2015), and Nosofsky and Zaki (1998) report that classification training led to superior learning compared to observation training. On the other hand, Levering and Kurtz (2015) and Patterson& Kurtz (2020) observed either no effect or even a preference for observation over classification (also see Corral & Carpenter, 2024).

Focus on Rule Adoption Following researchers such as Maddox et al. (2010) and Schnyer et al. (2009), we argue that differences between training methods may not always emerge in overall accuracy but in the types of strategies learners adopt. That is, classification and observation may lead to different rule-learning outcomes. To test this, participants were randomly assigned to either classification or observation training. We adapted the design from Experiment 4 of Medin et al. (1982), in which multiple categorization strategies were associated with varying levels of accuracy. There were four feature dimensions in the design. The optimal categorization rule was an XOR rule, which required judgments based on two features and could achieve 100% accuracy. In addition, two suboptimal uni-dimensional rules were also possible, each of which achieved 75% accuracy but required judgments on only one feature. Using a classifier model, we can examine which strategies (optimal, suboptimal, or others) participants adopt and whether there are significant differences between the classification and observation conditions.

Rule Consistency and Dynamics In addition to identifying which rule a participant adopts, we also examine how consistently that rule is applied across trials. This allows us to assess the stability of learning and whether switching between strategies contributes to the variability in training effects observed across studies. Some benefits of classification training, as those seen in the testing effect literature (e.g., Roediger & Karpicke, 2006), may emerge only over time due to better retention. Moreover, differences in across trials may reveal that participants improve later, even if they initially perform worse. (Ell et al., 2006)

Prediction Both the optimal XOR rule and the suboptimal uni-dimensional rules are verbalizable, rule-based structures (Ashby et al., 1998; Ashby & Maddox, 2005). Based on prior research, we predict that classification training will generally result in higher accuracy than observation training. However, as suggested by Carvalho and Goldstone (2015), Hsu and Griffiths (2010), and Levering and Kurtz (2015), observation training may promote participants to focus on commonalities within categories and be more sensitive to the relationships

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

between features. Consequently, participants in the observation condition may be more likely to discover the optimal XOR rule, which requires recognizing correlations between features. Thus, we hypothesize that the observation condition will lead to a higher proportion of participants adopting the optimal strategy and, therefore, achieving higher accuracy. Since these two explanations contradict each other, we do not have a clear prediction regarding the accuracy differences between the two training methods. Although overall accuracy differences may not be clear-cut, we expect training methods to produce distinct profiles in strategy adoption and consistency.

Methods

Participants

A total of 192 undergraduate students from Syracuse University participated in this experiment for course credit in an introductory psychology course. Participants were randomly assigned to either the observation or classification condition, with 96 participants in each condition.

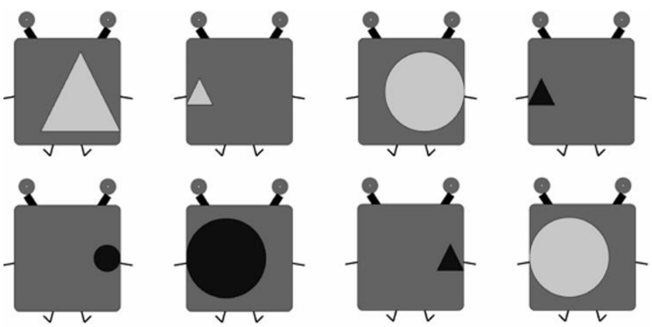

Figure 1: Examples of alien stimuli used in this study. The features are the shape, size, color, and position of the patch on their body.

Stimuli

The stimuli consisted of aliens with various patches on their bodies, as illustrated in Figure 1. Four dichotomous feature dimensions were used: patch size (large or small), patch color (dark or light), patch shape (circle or triangle), and patch position on the body (left or right). This resulted in a total of 16 unique aliens. The category structure followed the design from Medin et al. (1982), Experiment 4 (see Table 1). Aliens were divided into training and transfer stimuli, with participants learning from the training items and tested on both the training and transfer items. The optimal separation rule was an XOR rule that was based on the third and fourth dimensions. When the third and fourth dimension both had the values as 0s or 1s, the alien was in one category (Alkin); when two features had different values, the alien was in another category (Bafster). As a result, it required recognizing the relationship between two features. In addition, two suboptimal uni-dimensional rules, based on either the first or second dimension, yielded 75% accuracy in the training set. A critical aspect of the design was to examine whether participants could recognize the correlation between features (as the XOR rule) and whether they would adopt the optimal or suboptimal rules.

To control for potential salience effects of the features, the assignment of dimensions was counterbalanced across participants. Each participant was randomly assigned to one of the combinations of stimuli.

Table 1: Category Structure

Training set					Transfer set				
	Feature 1	Feature 2	Feature 3	Feature 4		Feature 1	Feature 2	Feature 3	Feature 4
A1	1	1	1	1	T1	0	0	0	0
A2	1	1	0	0	T2	0	0	1	1
A3	0	1	1	1	T3	0	1	0	0
A4	1	0	0	0	T4	1	0	1	1
B1	0	0	1	0	T5	1	1	1	0
B2	0	0	0	1	T6	1	1	0	1
B3	1	0	1	0	T7	0	1	1	0
B4	0	1	0	1	T8	1	0	0	1

Note. On the left side are the training items and on the right side are the transfer items that are only presented in tests. Each row represents a kind of alien with four dichotomous features. A1, A2, A3, and A4 are the Alkin exemplars while B1, B2, B3, and B4 are the Bafster exemplars.

Procedure

Participants were instructed to learn to categorize the aliens into two categories, "Alkin" and "Bafster," without prior knowledge of the categories. Each participant underwent training and testing phases, with the training format depending on the assigned condition.

In the classification condition, participants were presented with an alien and were asked to classify it by pressing the corresponding category key. After providing their response, participants received corrective feedback and were instructed to learn from the feedback. In the observation condition, the correct category label was presented with the alien, and participants were asked to study and learn the categorization. Participants were required to press the corresponding category key before proceeding to the next trial, ensuring greater alignment of responses across the two conditions. In both conditions, each trial involved the presentation of a single alien, and the training set consisted of 8 aliens. Participants completed 12 training blocks before moving to the test phase

An endorsement task was used in the test section to reduce the transfer appropriate processing effect. In each trial, participants were presented with an alien and a category label (either Alkin or Bafster). They were asked to judge whether the combination was correct (yes/no), without receiving feedback. All 16 aliens were tested, with 8 being old (from the training set) and 8 being new (transfer items). Each alien was presented twice, and the order of trials was randomized.

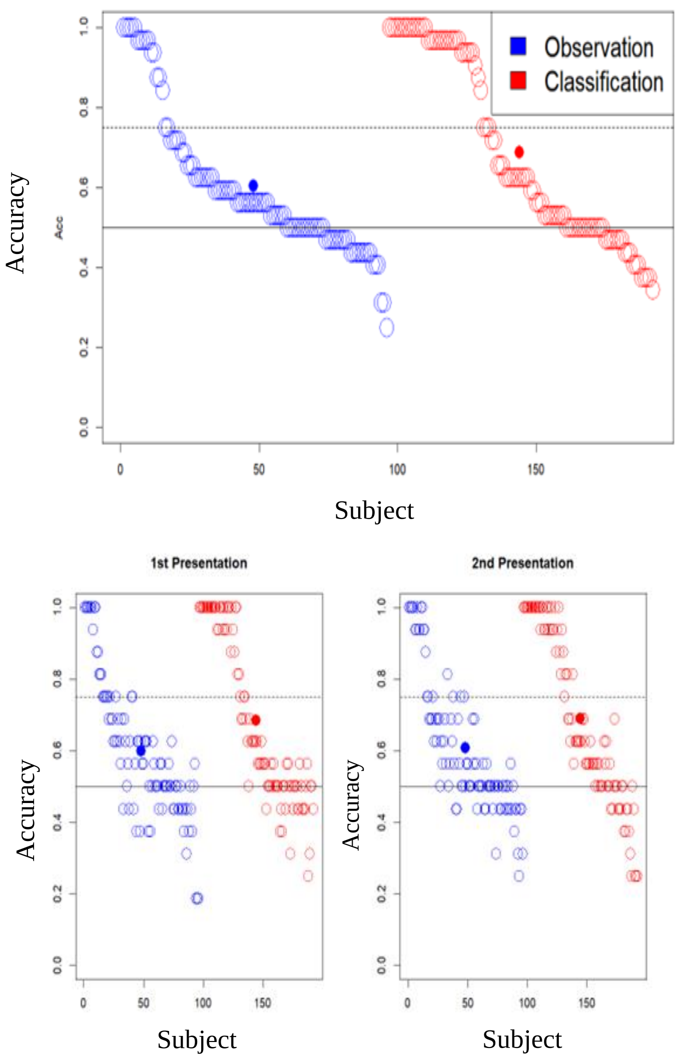

Figure 2: Proportion on correct responses in test. Open circles represent individual performance, and close circles represent the group average. The upper panel is the data cumulated from both first and second presentation while the lower two panels show the separate data. The solid lines are drawn on the random guessing criteria .50 and the dashed line are drawn on the suboptimal criteria .75.

Results

Training performance

In the classification condition, participants started with an average accuracy of 47% in the first training block and reached an average accuracy of over 80% in the final block. Seventy-two out of 96 participants (75%) achieved an accuracy equal to or greater than the suboptimal rule criterion. No judgment was requested in the observation condition, so no accuracy data were collected for that group. However, both conditions exhibited a decrease in reaction time across blocks, indicating learning progress.

Test performance

The main focus of this study was to compare performance between the classification and observation conditions during the test phase. Participants in the classification condition achieved an average accuracy of 69%, while those in the observation condition had an average accuracy of 60%. A *t*-test revealed that participants in the classification condition performed significantly better than those in the observation condition, $t(190) = 2.88$, $p < .01$. As shown in the upper panel in Figure 2, the classification condition had a higher proportion of participants achieving 100% accuracy and with fewer participants falling below chance.

To further investigate this effect, a 2 ×2 ×2 mixed-design ANOVA was conducted, with item type (old vs. new) and presentation (first vs. second) as within-subject factors, and condition (classification vs. observation) as a between-subjects factor. Significant main effects were found for item type ($F(1, 570) = 160.21$, $p < .001$) and condition ($F(1, 190) = 8.30$, $p = .004$). Participants showed higher accuracy on old items, and the classification condition showed a performance advantage over observation. The pattern was consistent across both presentations, with participants in the classification condition outperforming those in the observation condition. The interaction between item type and condition was also significant. Post hoc tests revealed that while classification outperformed observation on both old and new items, the difference was larger on old items. No other significant effects were observed.

Strategy identification

In addition to examining group-level accuracy, we explored whether different training methods influenced the strategies participants used to categorize the aliens. An example of this is shown in the lower panels in Figure 2, where the participants' order on the x-axis was kept consistent across both presentations. However, the data points for both the first and second presentations do not follow a monotonically decreasing function, suggesting that there was a shift in accuracy ranking within participants across presentations. Notably, the points for participants in the observation condition appeared more dispersed, indicating a larger variation in accuracy between the first and second presentations. To further investigate this, we conducted a strategy identification analysis for each participant across both presentations using a Bayesian classifier.

Model We used a hierarchical Bayesian model to identify strategies employed by participants in the task. The analysis was conducted using R (Ihaka & Gentleman, 1996) with the rjags package (Plummer et al., 2016).

The condition parameter, cond[i], was coded as 1 for participants in the classification condition and 0 for those in the observation condition. The model included a latent categorical variable, z[i], which assigned each participant to one of 11 possible rules, including the optimal XOR rule, two suboptimal rules (75% accuracy), other uni-dimensional rules, and random guessing. The probability distribution for z[i] was modeled as a categorical distribution, and we used a uniform Dirichlet prior on this distribution, assuming no strong preference for any particular rule.

An individual-specific parameter, theta[i], represented the probability of a participant classifying the aliens as "Alkin" if they were using the random guessing strategy (i.e., when z[i] = 11). This parameter followed a flat uniform prior between 0 and 1.

The probability that participant i would respond "Alkin" on item j was represented by prob[i,j]. Given z[i] = k, the probability prob[i,j] aligned with the corresponding rule k. Error probabilities, eps[i,j], were incorporated to account for cases where participants provided answers inconsistent with the rule. Thus, the probability of giving the correct answer was (1 - eps), and the probability of an incorrect answer was eps. We included error parameters for each condition and presentation (i.e., eps_cla1, eps_cla2, eps_obs1, and eps_obs2) to compare differences between the classification and observation conditions across both presentations. Prior distributions for these error parameters were Beta(1, 3), reflecting a belief that errors would be relatively infrequent.

Finally, the observed response, y[i,j], followed a Bernoulli distribution:

$$y[i,j] = dbern(prob[i,j])$$

The parameters z[i], theta[i], eps_cla, and eps_obs were estimated to best fit the observed responses.

Results The result of the strategy identification (z[i]) for both conditions are presented in Table 2 and 3. In the classification condition, participants' rule adoption was as follows in the first presentation: 27 adopted the 75% rule, 40 adopted the optimal 100% rule, 23 used random guessing, and 6 used other strategies. In the second presentation, these numbers shifted to 32, 42, 13, and 9, respectively. In the observation condition, 25 participants adopted the 75% rule, 36 adopted the 100% rule, 9 used random guessing, and 26 used other strategies in the first presentation. In the second presentation, the distribution changed to 29, 34, 20, and 13, respectively.

The most noticeable difference between conditions occurred in the first presentation of the observation condition, where a larger proportion of participants using the other uni-dimensional rule in the observation condition compared to the classification condition. A chi-square analysis confirmed that there were no significant differences between conditions in the second presentation(X^2 (1) = 13.03, $p < .001$), but the difference was significant in the first presentation(X^2 (1) = 3.20, $p = .36$). Moreover, the ratio between the first and second presentation was significantly different in the observation condition, but not in the classification condition. However, when comparing the adoption of the optimal (100%) rule versus the suboptimal (75%) rules, no significant differences between the two conditions were observed at either presentation.

Table 2: Cross Table for Strategy Identified in both presentations in Classification

Classification

1st \ 2nd	75% 1-D rules	100% XOR rule	Random	Other 1-D rules	
75% 1-D rules	24	0	3	0	27
100% XOR rule	1	38	0	1	40
Random	5	4	9	5	23
Other 1-D rules	2	0	1	3	6
	32	42	13	9	

Note. The grey numbers are the sum in either presentation.

Table 3: Cross Table for Strategies Identified in both presentations in Observation

Observation

1st \ 2nd	75% 1-D rules	100% XOR rule	Random	Other 1-D rules	
75% 1-D rules	13	8	3	1	25
100% XOR rule	6	24	5	1	36
Random	4	0	3	2	9
Other 1-D rules	6	2	9	9	26
	29	34	20	13	

Note. The grey numbers are the sum in either presentation.

Further analysis grouped participants by whether they maintained the same strategy across both presentations or switched strategies. In the classification condition, 74 out of 96 participants used the same strategy across both presentations, compared to only 49 out of 96 in the observation condition. A chi-square test revealed a significant difference in the proportion of participants who maintained their strategies between the two conditions (X^2 (1) = 13.03, $p < .001$).

The estimated error parameters for the classification condition were .13 for both the first and second presentations, whereas in the observation condition, the error rates were .22 and .19, respectively. These results suggest that participants in the classification condition were better at adhering to the strategies they adopted. To quantify this further, we calculated the average match proportion, which represents the consistency between participants' responses and the strategies that they were identified as using. Results, shown in Table 4, indicated that, across both presentations and for both old and new items, the classification condition consistently demonstrated lower or equivalent error rates compared to the observation condition. Aggregating across all rule types, we calculated the deviation between each participant's responses and the idealized response pattern of the rule they were identified as using. As shown in Figure 3, participants in the observation condition exhibited greater deviations from their identified rule across both old/new items and first/second presentations. This suggests that their responses were less consistent with the rule they were assumed to follow, compared to participants in the classification condition.

Finally, the estimated guessing rate parameters (theta[i]) were close to .50 for all participants and did not significantly differ between conditions, suggesting no systematic bias toward one category label or the other.

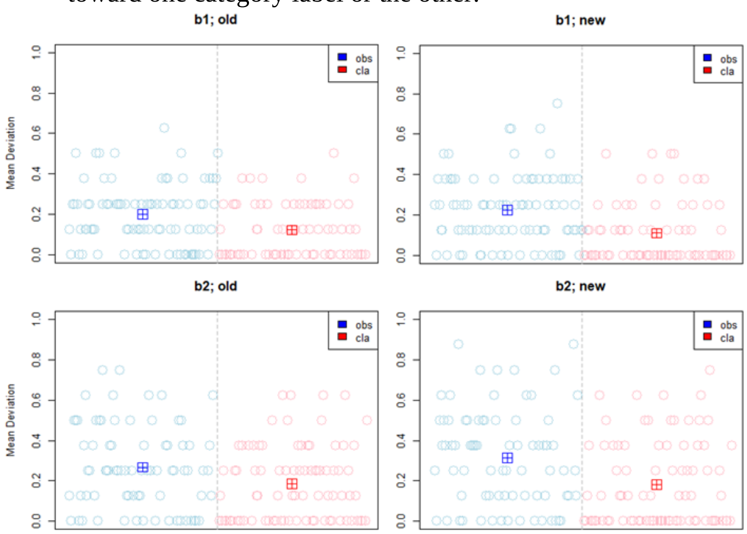

Figure 3: Deviation from the rule. Open circles represent individual deviation scores, and boxes represent the group average. The upper right panel is the data from new items in the first presentation; the upper left panel is the data from old items in the first presentation. The lower right panel is the data from new items in the second presentation; the lower left panel is the data from old items in the second presentation.

Discussion

This study differs from previous literature by directly comparing classification and observation training in a categorization task that allowed participants to adopt various strategies. We found that participants in the classification condition outperformed those in the observation condition in terms of average accuracy during the test phase. Through Bayesian classifier analysis, we identified the strategies adopted by participants, and the results suggest that the advantage of classification may stem from more participants adopting the optimal rule and demonstrating greater consistency in following their chosen rule, whether optimal or suboptimal. Additionally, participants in the classification condition were less likely to switch strategies compared to those in the observation condition

In our task, both the optimal rule and suboptimal rules followed a rule-based structure. Unlike previous studies (e.g., Levering & Kurtz, 2015; Patterson& Kurtz, 2020), which predicted observation training would lead to better categorization performance, our findings suggest that classification training may offer a distinct advantage. This contradicts previous studies (e.g., Carvalho & Goldstone, 2015; Hus & Griffiths, 2010; and Levering & Kurtz, 2015), which suggest that observation training would promote optimal rule adoption by increasing sensitivity to within-category structures. Our results imply that observation training might not necessarily facilitate the acquisition of the optimal XOR rule over other, potentially suboptimal, strategies—or that the benefits of classification outweigh those of observation.

Table 4: The Estimation of Error Parameters.

	75% rule 1	75% rule 2	100% XOR rule		75% rule 1	75% rule 2	100% XOR rule
Obs_p1_Old	0.78	0.81	0.84	Obs_p2_Old	0.67	0.68	0.78
Cls_p1_Old	0.78	0.82	0.95	Cls_p2_Old	0.72	0.76	0.91
Obs_p1_New	0.80	0.80	0.75	Obs_p2_New	0.66	0.64	0.68
Cls_p1_New	0.89	0.88	0.91	Cls_p2_New	0.81	0.71	0.89

Note. Obs represents the observation and Cls represents the classification. P1 represents the first presentation and P2 represents the second presentation. Old represents the results on the old items (training items) and New represents the results on the new items (transfer items).

Explanations of the Difference between Classification and Observation

Commitment to Strategy One key difference between classification and observation is the level of commitment to the learned strategy. Our results show that participants in the classification condition were more committed to their adopted strategies, and they were less likely to deviate from them. This suggests that classification training encourages stronger strategy commitment. As classification requires participants to actively generate a response, they cannot remain uncertain about categorization, unlike participants in the observation condition. This active process forces participants to update their beliefs quickly and make decisions that exceed a threshold for response. Consequently, classification participants develop more robust posteriors, leading to more committed and consistent strategy use.

Chang et al. (2024) introduced the "extra learning opportunity hypothesis" to explain the differences between classification and observation. They proposed that, in each classification trial, participants learn not only from feedback but also from their own response prior to feedback (i.e., the extra learning opportunity). When implemented in a revised version of the Locally Bayesian Learning model (LBL, Kruschke, 2006), this hypothesis helped explain the differential learning outcomes between the two conditions. Although this hypothesis is distinct from the *"one-answer challenge"* explanation we propose here, both predict that classification fosters stronger strategy knowledge due to more frequent updating. We argue that generally participants who are more committed to their strategies perform better, as they are less likely to make errors. But overcommitting to lure rules may lead to low performance. This suggests that classification training is most beneficial when the optimal rule is clear and easily identifiable, while observation training may be more useful when suboptimal rules are distracting or inaccurate.

Gradual Learning Another possible explanation for the differences between training methods is that participants in the observation condition may learn more slowly. It is plausible that participants in the observation condition were less prepared at the start of the test phase, but gradually caught up as they continued testing (e.g., learning during test, LDT; Hu & Nosofsky, 2022; Nosofsky & Zaki, 2004). Evidence for this comes from the fact that the classification advantage was only observed during the first presentation of test items, while the second presentation showed no such difference. Additionally, error rates decreased from the first to the second presentation. This suggests that observation training may be more effective in tasks that are easier or when learning is massive. Future research could explore this possibility by extending the test phase and examining performance dynamics over a greater number of trials.

Retrieval Practice

Classification training, which requires participants to actively generate a response, engages more retrieval processes than observation training. According to principles of retrieval practice (Carrier & Pashler, 1992), such active retrieval can enhance performance. Consistent with this, our results showed that participants in the classification condition outperformed those in the observation condition, particularly on previously seen (old) items. This supports the idea that retrieval practice contributes to the superior performance observed with classification training.

Dual-system view of Categorization The dual-system model of categorization, specifically the Competition between Verbal and Implicit Systems (COVIS) proposed by Ashby et al. (1998), posits two learning systems: a verbal system that operates through rule-based, verbalizable knowledge, and an implicit system that learns through procedural associations. Ashby et al. (2002) found that classification training outperformed observation in learning non-verbalizable (information-integration, II) structures, while the two methods yielded little difference results for rule-based structures. These findings have been used to support the dual-system perspective, although the view has faced criticism (e.g., Donkin et al., 2015; Dunn et al., 2012).

In our study, we observed a similar pattern—classification led to better learning overall, and the difference between training methods diminished over time. However, since we focused solely on rule-based structures (i.e., uni-dimensional and XOR rules), our results do not directly address the distinction between verbal and implicit systems. Thus, while our findings align with some predictions of the dual-system model, they do not speak to the broader debate on the existence of multiple categorization systems. Future research could extend this work by examining rule adoption and consistency in tasks involving non-verbalizable category structures.

Strategy Adoption and Strategy Consistency as a Measure of Difference

Finally, our findings challenge the notion that overall accuracy is the most reliable indicator of categorization performance. Group-level accuracy reflects a mixture of various factors, including strategy adoption, response consistency, certainty, and stability. As shown in this study, decomposing accuracy into individual-level indicators—such as the proportion of participants using different strategies, their commitment to those strategies, and the probability of errors—provides a more nuanced evaluation of performance. Focusing exclusively on average accuracy may obscure underlying differences in how participants engage with the task.

Conclusion

In summary, this study aimed to explore the differences between classification and observation training in categorization tasks. The results indicate that classification training leads to better performance in tasks that involve multiple accountable rules. Specifically, classification participants showed stronger commitment to their strategies, which resulted in less strategy-switching and fewer errors, enhancing their accuracy. These findings contribute to a better understanding of the inconsistent results observed in categorization research and offer new insights into how different training methods influence strategy adoption and performance.

References

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review*, *105*(3), 442.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.*, *56*(1), 149-178.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. Memory & cognition, 30(5), 666-677.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & cognition*, *20*, 633-642.
Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic bulletin & review*, *22*, 281-288.
Chang, Y. W., Aytac, S., Mendoza Gonzalez, C. G., Kalish, M., & Corral, D. (2024). Extending the Locally Bayesian Learning Model to Exemplar-Based Categorization with Continuous Features. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).
Corral, D., & Carpenter, S. K. (2024). Acquiring complex concepts through testing and explanatory feedback. *Cognitive Research: Principles and Implications*, *9*, 1–22.
Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: a case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 933.
Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: the limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 840.
Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly journal of experimental psychology*, *68*(6), 1203-1222.
Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *Quarterly Journal of Experimental Psychology*, *65*(8), 1537-1562.
Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. *Neuropsychologia*, *44*(10), 1737-1751.
Hsu, A., & Griffiths, T. (2010). Effects of generative and discriminative learning on use of category variability. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).
Hu, M., & Nosofsky, R. M. (2022). Exemplar-model account of categorization and recognition when training instances never repeat. *Journal of experimental psychology: Learning, memory, and cognition*, *48*(12), 1947.
Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, *5*(3), 299-314.
Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological review*, *113*(4), 677.
Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & cognition*, *43*, 266-282.
Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Rule-based and information-integration category learning in normal aging. *Neuropsychologia*, *48*(10), 2998-3008.
Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(1), 37
Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*(4), 247-255
Patterson, J. D., & Kurtz, K. J. (2020). Comparison-based learning of relational categories (you'll never guess). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 851.
Plummer, M., Stukalov, A., & Denwood, M. (2016). rjags: Bayesian graphical models using MCMC. *R package version*, *4*(6), 4.
Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological science, 17(3), 249-255.
Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. *Neuropsychologia*, *47*(13), 2995-3006.
Zaki, S. R., & Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory & Cognition*, *32*(3), 390-398.