

Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2002)

F. GREGORY ASHBY and SHAWN W. ELL
University of California, Santa Barbara, California

Nosofsky and Kruschke (2002) argued that the single-system ALCOVE model (Kruschke, 1992) can account for the dual-task category learning data reported by Waldron and Ashby (2001). In our reply, we argue that Nosofsky and Kruschke overstated the ability of ALCOVE to account for the Waldron and Ashby results. In fact, ALCOVE has difficulty with these data, and we show that the only versions of ALCOVE that actually fit the Waldron and Ashby accuracy data make incorrect predictions about other previously unreported features of that experiment. We also show that the dual-system COVIS model (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) naturally predicts these results.

Waldron and Ashby (2001) reported the results of an experiment in which participants learned simple and complex category structures under typical single-task conditions and when performing a simultaneous numerical Stroop task. In the simple categorization tasks, each set of contrasting categories was separated by a unidimensional explicit rule, whereas the complex tasks required the integration of information from three stimulus dimensions and resulted in rules that were difficult to verbalize. The concurrent Stroop task dramatically impaired learning of the simple structures but did not significantly delay learning of the complex structures. Waldron and Ashby argued that single-system accounts of category learning naturally predict that a dual task should interfere more strongly with the complex task than with the simple task (it should be easier to perform two simple tasks at the same time than two difficult tasks). Since the opposite pattern was observed, Waldron and Ashby argued that their results provided evidence of at least two separate category learning systems.

In response, Nosofsky and Kruschke (2002) reported the results of a large computational study that investigated the predictions of Kruschke's (1992) single-system ALCOVE model of category learning in the Waldron and Ashby (2001) experiment. On the basis of these investigations, Nosofsky and Kruschke argued that Waldron and Ashby's dual-task results do not pose a serious challenge for single-system models of category learning.

We believe that Nosofsky and Kruschke overstated the ability of ALCOVE to account for the Waldron and Ashby (2001) results. In fact, ALCOVE has difficulty with these data, and we show that the only versions of ALCOVE that actually fit the Waldron and Ashby accuracy data make incorrect predictions about other previously unreported features of that experiment.

ALCOVE Has Difficulty Fitting the Waldron and Ashby (2001) Data

Nosofsky and Kruschke (2002) focused on the question of whether ALCOVE can account for the ordinal result that the concurrent task used by Waldron and Ashby (2001) interfered more strongly with the learning of the simple category structures than with the learning of the complex structures. Their comment ignores the more ambitious question of whether ALCOVE can fit the Waldron and Ashby data *quantitatively*. In fact, despite having more free parameters (i.e., five) than there are datapoints (i.e., four), ALCOVE has difficulty providing accurate fits to the Waldron and Ashby data.

First, consider just the single-task control data. The Waldron and Ashby (2001) participants reached criterion in this condition in about 15 trials with the simple structures and in about 50 trials with the complex structures. These data therefore support the intuition that it is much easier for people to learn categories that are separated by a simple unidimensional rule than categories that are separated by a complex three-dimensional rule that is difficult to describe verbally. Waldron and Ashby argued that this difference in difficulty occurs because people use an explicit, logical-reasoning system to learn the simple unidimensional rules but some other implicit-like system to learn the complex rules. Since ALCOVE is a single-system model, it obviously must learn both types of category structures with the same cognitive architecture. An immediate challenge for any single-system model is to account for the observed difference in difficulty in the control data, which is much larger than would be predicted, say, by an ideal observer.¹

In fact, the simulations reported by Nosofsky and Kruschke (2002) indicate that ALCOVE often underpredicts this difficulty difference. This can be seen in Nosofsky and Kruschke's Figure 2, which shows ALCOVE predictions for 140 different combinations of the ALCOVE parameters. Consider the predictions of ALCOVE when there is no attentional learning (i.e., $\lambda_a = 0$). These predictions are generated by randomly selecting attention weights for the various stimulus dimensions and then holding these weights fixed throughout the simulated experiment. As such, they illustrate the predictions of ALCOVE when attention is randomly distributed among the dimensions. Note that ALCOVE does predict that the three-dimensional task is more difficult than the unidimensional task, but also

Preparation of this article was supported by Grant BCS99-75037 from the National Science Foundation. We thank Elliott Waldron and Todd Maddox for their helpful suggestions in preparing this article. Correspondence concerning this article should be addressed to F. G. Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106 (e-mail: ashby@psych.ucsb.edu).

note that for all 20 combinations of the other parameters, ALCOVE underpredicts the observed difficulty difference.

When attentional learning occurs, ALCOVE can account for the observed difference between simple and complex category learning, but in almost all such cases, it predicts overly poor performance in the simple unidimensional condition. A careful examination indicates that in almost all of the 140 cases shown in Nosofsky and Kruschke's (2002) Figure 2, ALCOVE either predicts overly poor performance with the simple structures or too little difference between the simple and complex structures.

ALCOVE comes close to a perfect fit of the Waldron and Ashby (2001) control data when $\lambda_w = 0.07$, $c = 1.50$, and the attention learning rate $\lambda_a = 0.160$ (and $\phi = 4.5$). When the attention learning rate is set to 0, these same parameter values allow ALCOVE to also provide good fits to the dual-task data of Waldron and Ashby. However, note that with any value of the attention learning rate greater than 0, the model seriously underpredicts the amount of interference that the dual task causes in learning the simple category structures. In fact, this result generalizes to other sets of parameter values. For any set of parameters that allow ALCOVE to roughly fit the control data, reasonable fits to the dual-task data require setting $\lambda_a = 0$.

Thus, the Nosofsky and Kruschke (2002) simulations show that ALCOVE can fit the Waldron and Ashby (2001) data, but only under the extreme assumption that there is no attentional learning at all in the dual-task conditions. This version of ALCOVE makes a very strong prediction. Specifically, since the attention weights are originally set to random values, ALCOVE predicts that after reaching criterion accuracy on the simple-category structures, observers will have no idea that only one dimension was relevant in the dual-task condition.

To test this prediction of ALCOVE we ran 5 more participants in the Waldron and Ashby (2001) experiment. The present experiment differed from the Waldron and Ashby experiment in two respects. First, the two experimental sessions from the Waldron and Ashby experiment were condensed into a single session. Each participant first learned two simple (i.e., unidimensional) and two complex (i.e., three-dimensional) rules under standard single-task conditions (with rule order being random across participants). Second, each participant learned a single, new unidimensional rule under dual-task conditions. After achieving criterion accuracy (eight consecutive correct responses) in the dual-task condition, the participants were asked to describe the strategy that they used in the categorization portion of that part of the experiment.

The accuracy results from this experiment are shown in Figure 1. Note that they are similar to the results of Waldron and Ashby (2001; although the difference between the results for the simple and complex structures in the control condition is even more pronounced in Figure 1 than in Waldron & Ashby). Even so, 4 of the 5 participants perfectly described the unidimensional rule that they used in the dual-task condition. The 5th participant described a

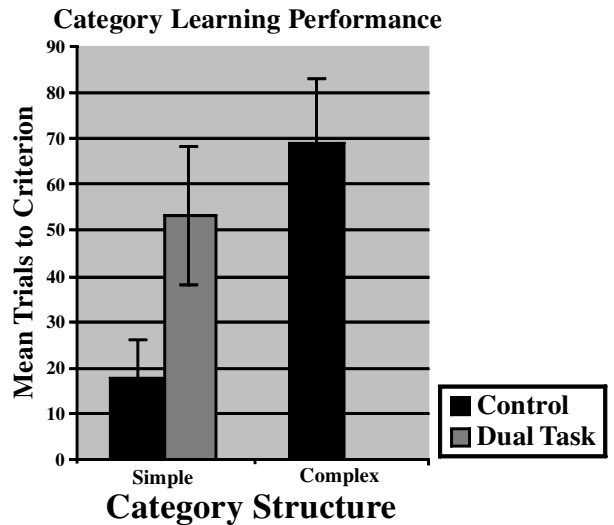


Figure 1. Mean number of trials to criterion for the control and dual-task data.

(two-dimensional) conjunctive rule that included the critical stimulus dimension. Thus, in contrast to the ALCOVE predictions, every participant showed significant attentional learning, and 4 of the 5 participants showed perfect attentional learning.

One possibility not ruled out by our results, so far, is that ALCOVE might be able to roughly fit the Waldron and Ashby (2001) data with some small, but nonzero value of the attentional learning rate (i.e., λ_a). As long as the attentional learning rate is greater than zero, some attentional learning will occur, so it is of interest to ask whether enough learning will occur in such cases to explain the accuracy of our verbal reports. The smallest nonzero value of λ_a investigated by Nosofsky and Kruschke (2002) was $\lambda_a = 0.005$. Figure 2 of Nosofsky and Kruschke shows that none of the fits with this value of the attentional learning rate were good, but better results might be obtained with even smaller values of λ_a . However, even with $\lambda_a = 0.005$, ALCOVE predicts very little attentional learning over the course of the experiment. For example, we conducted a series of simulations of ALCOVE with $\lambda_a = 0.005$. Like Nosofsky and Kruschke, we fixed ϕ at 4.5, let c range from 0.75 to 4.5, and λ_w range from 0.01 to 0.10. In each case, we simulated the performance of ALCOVE with the simple category structures until criterion accuracy was reached (eight consecutive correct responses). At the end of this learning period, we examined the final values of the attention weights. The results are shown in Figure 2. Note that in every case, ALCOVE allocated less than 40% of the total attention to the single relevant dimension. Of the three irrelevant dimensions, they each received at least 19.8% of the total attention. Thus, even with an attentional learning rate that is too large to allow ALCOVE to provide good fits to the Waldron and Ashby data, ALCOVE still predicts that at the end of training, observers will be allocat-

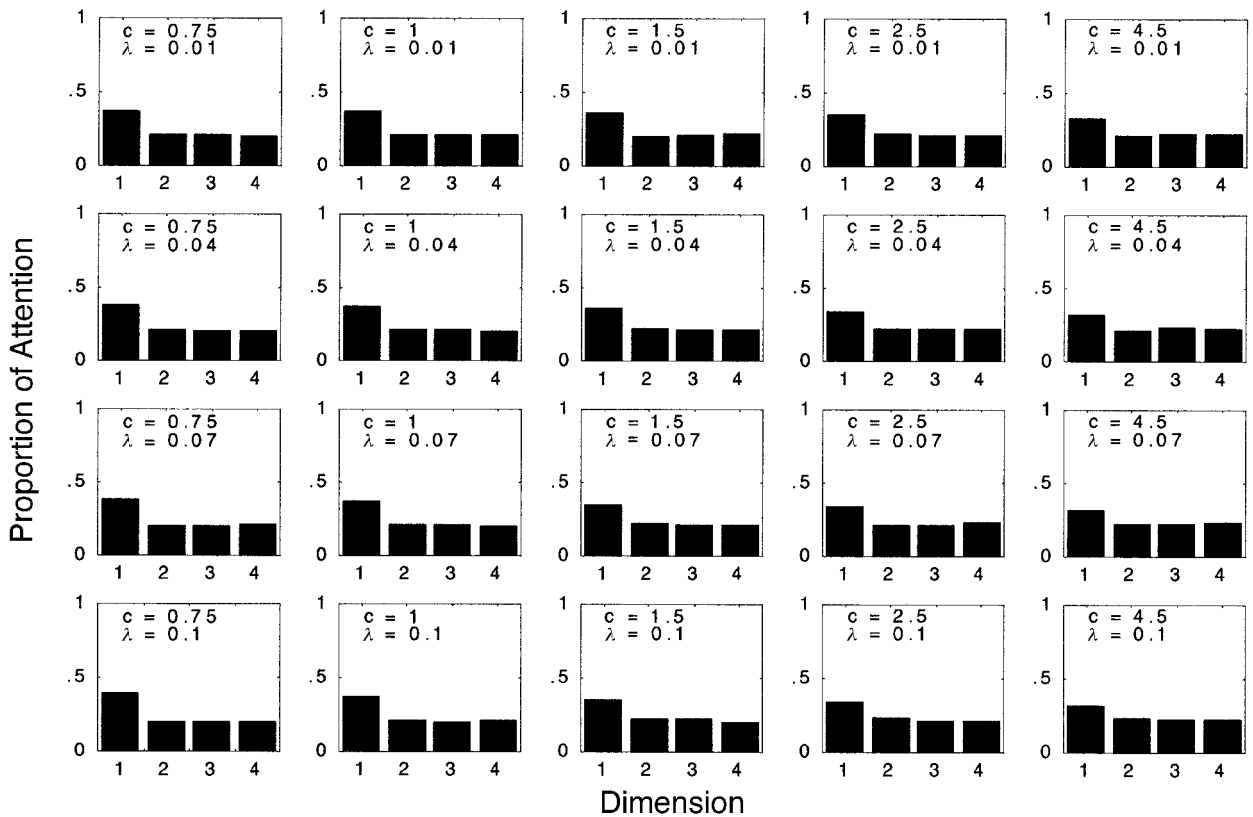


Figure 2. Proportion of attention allocated to each of the four stimulus dimensions averaged across 1,000 replications. The value of c varied from 0.75 to 4.5 (across columns in the figure) and λ_a varied from 0.01 to 0.1 (across rows in the figure), whereas λ_w and ϕ were held constant across at .005 and 4.5, respectively.

ing most of their attention to irrelevant dimensions. This prediction was not supported by our data.²

COVIS Easily Accounts for the Waldron and Ashby (2001) Results

A second interesting and important question is whether the dual-system model COVIS (Ashby et al., 1998) can account for these verbal reports. COVIS assumes separate explicit (rule-based) and implicit (procedural learning-based) category learning systems that compete throughout training. A system weight determines the relative contribution that each system makes to the observable categorization response. Initially, the system weight favors the explicit system, but it is then adjusted up or down, depending on the relative success of the two systems.

Waldron and Ashby (2001) showed that COVIS fit their data quantitatively, but they did not address the question of whether the best-fitting version of COVIS predicted that, by the end of training, the participants would be aware that a single dimension was relevant with the simple category structures. In the COVIS explicit system, the process of selecting among alternative explicit categorization rules is mediated by frontal cortical structures. Stroop tasks are known to activate frontal cortex (Bench et al., 1993), so

Waldron and Ashby hypothesized that the effect of the numerical Stroop dual task would be to decrease the efficacy of the explicit system to select new categorization rules.

COVIS correctly predicts the verbal reports that we collected only if two conditions are met. First, the system weight on the explicit system must be reasonably large in order to guarantee that a significant percentage of responses were generated by the explicit system. Second, the explicit system must learn the optimal rule. To investigate these issues we conducted a series of simulations of COVIS.

As mentioned above, Waldron and Ashby (2001) hypothesized that the concurrent task would decrease the ability of the explicit system to select new categorization rules. In COVIS, this ability is instantiated in the selection parameter λ (not to be confused with the ALCOVE learning rates λ_a and λ_w). Specifically, Waldron and Ashby hypothesized that the concurrent task would decrease λ . As in Waldron and Ashby, in our simulations of COVIS, we first fit the control data by crudely adjusting all parameters.³ Next, with all parameters except λ fixed at these values, we found the value of λ that best fit the dual-task data. Figure 3 shows that an excellent fit to the Waldron and Ashby accuracy data occurs when $\lambda = 1.55$ for the control data and $\lambda = 0.35$ for the concurrent data ($r^2 = .98$).

Waldron & Ashby (2001) Data

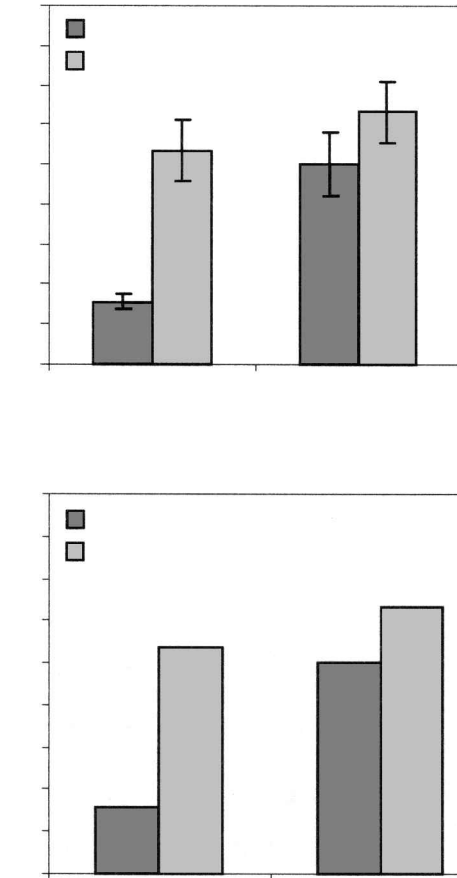


Figure 3. Waldron and Ashby (2001) data and COVIS predictions.

As mentioned above, our simulations focused on two questions. First, does the explicit system contribute to category learning under dual-task conditions? Second, does the explicit system learn the correct unidimensional rule in the presence of the dual task? Figure 4 shows values of the COVIS system weight after criterion accuracy is met for many different values of λ . Figure 4 clearly shows that for all values of λ , the system weight favors the explicit system. Thus, in all these cases, COVIS gives greater weight to the explicit system than to the implicit system.

On each trial, the COVIS explicit system selects the categorization rule with the highest salience, where saliences are updated after each trial according to past reinforcement history, the participant's tendency to perseverate, and the ability of the participant to select new rules (see Ashby et al., 1998, for details of this updating process). Figure 5 shows the saliences of the four unidimensional rules on each of the last eight trials before criterion accuracy was met with the unidimensional category structures. Note that for all values of λ , the optimal rule has the highest salience on virtually all trials. Thus, for this

entire range of λ , the COVIS explicit system is consistently using the optimal unidimensional decision rule at the end of training.

In summary, our simulations of the dual-system model COVIS produced several important results. First, COVIS provided excellent fits to the Waldron and Ashby (2001) accuracy data. Second, over a wide variety of parameter settings, including those that provided the best fits, the COVIS explicit system was more heavily weighted than was the implicit system, and the explicit system always learned the correct unidimensional decision rule by the end of training. Because of the last two results, COVIS correctly predicts that participants will be able to describe the optimal categorization rule for the simple category structures under the dual-task conditions.

Conclusions

Nosofsky and Kruschke (2002) showed that many combinations of ALCOVE's parameters allowed the model to account for the result that the concurrent task used by Waldron and Ashby (2001) interfered more strongly with the learning of the simple-category structures than with the complex structures. In our reply, we asked whether ALCOVE could fit the Waldron and Ashby data quantitatively. We showed that the only versions of ALCOVE that could fit the Waldron and Ashby data made the strong prediction that after reaching criterion accuracy on the simple category structures, participants would have no idea that only one dimension was relevant in the dual-task conditions. We reported new empirical evidence that

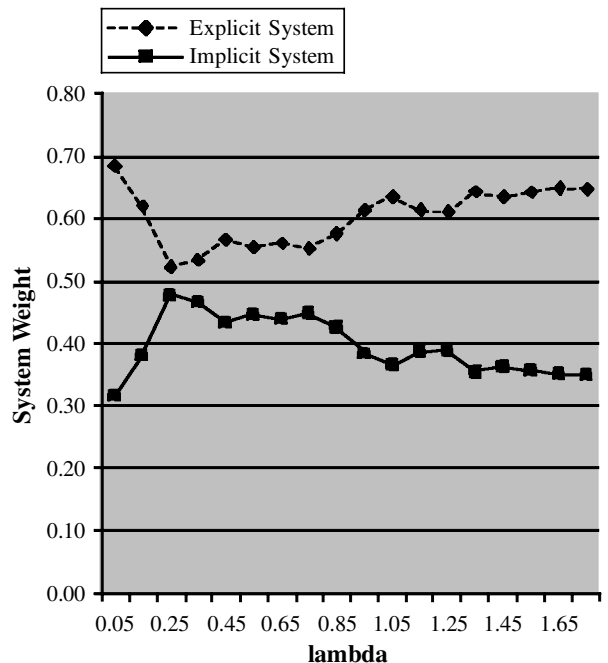


Figure 4. Average asymptotic system weights for the explicit (dashed line) and implicit (solid line) systems of COVIS as a function of changes in the selection parameter.

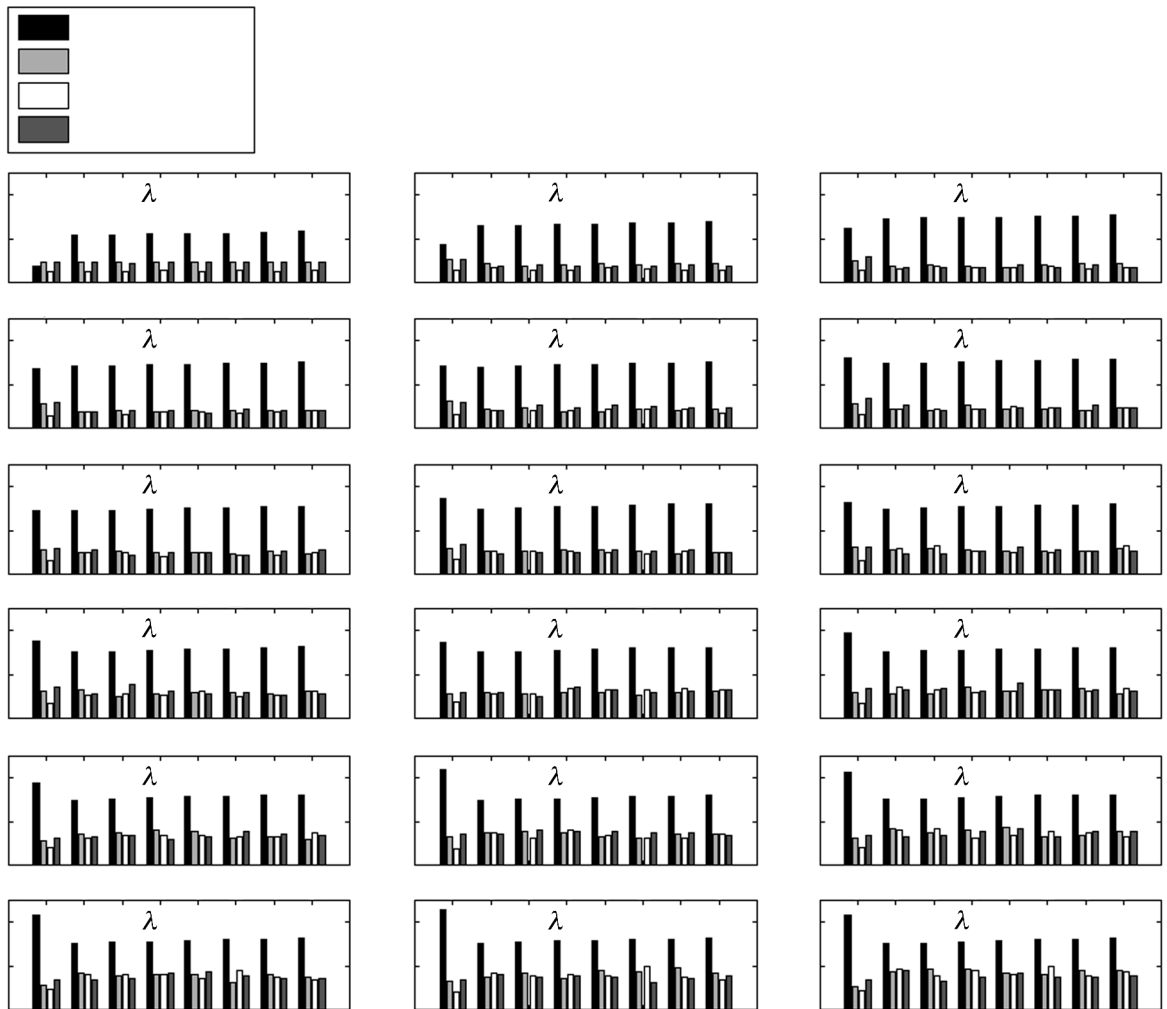


Figure 5. Salience in the COVIS explicit system of each of the four unidimensional explicit rules during the last eight correct responses. The data in each panel are an average across 200 replications for a different value of the selection parameter, λ . The COVIS explicit system always chooses the rule with the highest salience, indicating that in virtually every case, the COVIS explicit system was using the optimal rule.

strongly disconfirms this prediction, and we showed that the dual-system model COVIS naturally predicts these new data.

Although we believe that the results presented here are strong evidence that ALCOVE is not able to account for results in the Waldron and Ashby (2001) experiment, this does not mean that all single-system models of category learning are incompatible with the Waldron and Ashby data. In fact, we suspect that a clever researcher could construct such a model. For example, it might be possible to modify ALCOVE, so that it can account for the new verbal report data described here. For this reason, we believe that significant progress will be made toward answering the question of whether there are one or more category

learning systems only when a converging operations approach is used in which many different data sets from a variety of disparate experimental paradigms are simultaneously considered (e.g., Ashby & Ell, 2002).

REFERENCES

- ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A. U., & WALDRON, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.
- ASHBY, F. G., & ELL, S. W. (2002). Single versus multiple systems of learning and memory. In H. Pashler (Ed.) & J. Wixted (Vol. Ed.), *Stevens' Handbook of experimental psychology: Vol. 4 Methodology in experimental psychology* (3rd ed., pp. 655-692). New York: Wiley.
- BENCH, C. J., FRITH, C. D., GRASBY, P. M., FRISTON, K. J., PAULESU, E., FRACKOWAİK, R. S. J., & DOLAN, R. J. (1993). Investigations of the

- functional anatomy of attention using the Stroop test. *Neuropsychologia*, **33**, 907-922.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- NOSOFSKY, R. M., & KRUSCHKE, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, **9**, 169-174.
- WALDRON, E. M., & ASHBY, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, **8**, 168-176.

NOTES

1. An ideal observer would perform more poorly with the complex categories (assuming some noise). Both types of category structure are linearly separable, and in both cases, perfect accuracy can be achieved by responding with the category that has the most similar prototype. The category prototypes (i.e., centroids) are two units apart with the simple-category structures and $\sqrt{3}$ units apart with the complex structures. The difference between these two distances is much smaller, however, than the observed difficulty difference (i.e., 15 vs. 50 trials to criterion).
2. One could argue that if we had asked participants which dimension was most important, they would have been able to respond cor-

rectly, even if most of their attention was allocated to irrelevant dimensions. However, they were instructed to describe the strategy that they used during categorization. If most of the attention was allocated to irrelevant dimensions, it makes sense that at least some of the observers would mention this.

3. This process yielded the following numerical values: $\lambda = 1.55$, the perseveration parameter was set to 1.25, the increment on the salience of explicit rules following a correct response and the decrement on the salience of explicit rules following an incorrect response were both set to 0.02, the initial weights of the four alternative dimensional rules were each set to 0.25, the initial learning rate on the implicit system was set to .0005, and the initial system weights were set to .99 for the explicit system and .01 for the implicit system. The momentum and decay terms, and the initial learning rate for the system output node were irrelevant in these simulations owing to the use of binary-valued stimulus dimensions. See Ashby et al. (1998) for more details on the COVIS model.

(Manuscript received March 26, 2001;
revision accepted for publication December 17, 2001.)