Memory for exemplars in category learning

C. E. R. Edmunds (ceredmunds@gmail.com)
School of Psychology, Plymouth University
Plymouth, PL4 8AA, UK

Andy J. Wills (andy.wills@plymouth.ac.uk)
School of Psychology, Plymouth University
Plymouth, PL4 8AA, UK

Fraser N. Milton (F.N.Milton@exeter.ac.uk)
School of Psychology, University of Exeter
Exeter, EX4 4QG, UK

Abstract

Some argue that category learning is mediated by two competing learning systems: one explicit, one implicit (Ashby et al., 1998). These systems are hypothesised to be responsible for learning rule-based and information-integration category structures respectively. However, little experimental work has directly investigated whether people are conscious of category knowledge supposedly learned by the implicit system. Here we report one experiment that directly compared explicit recognition memory for exemplars between these two category structures. Contrary to the predictions of the dual-systems approach, we found preliminary evidence of superior exemplar memory after information-integration category learning compared to rule-based learning. This result is consistent with the hypothesis that participants learn information-integration category structures by using complex rules.

Keywords: category learning; memory; dual-systems; recognition;

One approach to categorization assumes that generalisation from past experiences to novel ones is mediated by two competing systems: one explicit and one implicit. The COVIS (COmpetition between Verbal and Implicit Systems) model is a popular instantiation of this approach (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby, Paul, & Maddox, 2011). In this model, the explicit, verbal system is described as learning by testing hypotheses using working memory. This system is assumed to optimally learn rule-based category category structures that can be easily verbalised such as the unidimensional structure shown in Figure 1A. In contrast, the implicit system is described as using procedural learning to associate areas of stimulus space with a motor response. This system is assumed to optimally learn category structures that cannot be easily verbalised such as the information-integration category structure shown in Figure 1B. Critically, the implicit system is assumed to "produce category knowledge that is opaque to declarative consciousness" (p.1, Smith et al., 2015).

Whether COVIS, and more broadly a dual-systems approach, adequately explains the processes of category learning is still a matter for debate. Proponents of COVIS argue that the case is closed, that the evidence for dual-systems approaches is overwhelming and that the field should move on to more interesting questions, such as those concerning the exact nature of the systems and how they interact with each other (Ashby & Maddox, 2011). In support of this view there is a large quantity of evidence that has been used to support COVIS (for a review see Ashby & Maddox, 2011). However, much of this evidence has been questioned (Edmunds, Milton, & Wills, 2015; Newell, Dunn, & Kalish, 2011; Stanton & Nosofsky, 2007; Zaki & Kleinschmidt, 2014). Also, despite the volume of studies, there is very little experimental work that directly investigates the key theoretical assumption that the learning of the implicit system is not available to consciousness. Instead, the focus has been on demonstrating that information-integration category structures are learned procedurally or demonstrating that learning of rule-based and information-integration categories are dissociable using experimental, neuropsychological or neuroscientific methods (Ashby & Maddox, 2005, 2011; Price, Filoteo, & Maddox, 2009). Therefore, the claim that the case is closed may be premature.

In the current study, we directly examine whether participants have conscious access to information about the information-integration categories they have learned. COVIS predicts that they do not, but some recent behavioral and neuroimaging evidence suggests otherwise. Behaviorally, Edmunds et al. (2015) found that the vast majority of participants were able to report a clear explicit strategy after training, regardless of whether they had been learning an information-integration or rule-based category structure. This was despite having met the criteria usually used in the COVIS literature to check that participants in the information-integration category structure condition are using the implicit system. This check uses a model-based strategy analysis inspired by General Recognition Theory (GRT; Ashby & Gott, 1988), a multidimensional version of signal detection theory. The failure of the model-based strategy analysis here may be because its output depends strongly on the set of strategies the modeller chooses to use, with the estimated proportion of "implicit" responders reducing substantially if more complex rule-based models are included (Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015).

Turning to neuroimaging evidence, in a recent study from our lab we found greater activation of the medial temporal lobe in information-integration category learning, rela-

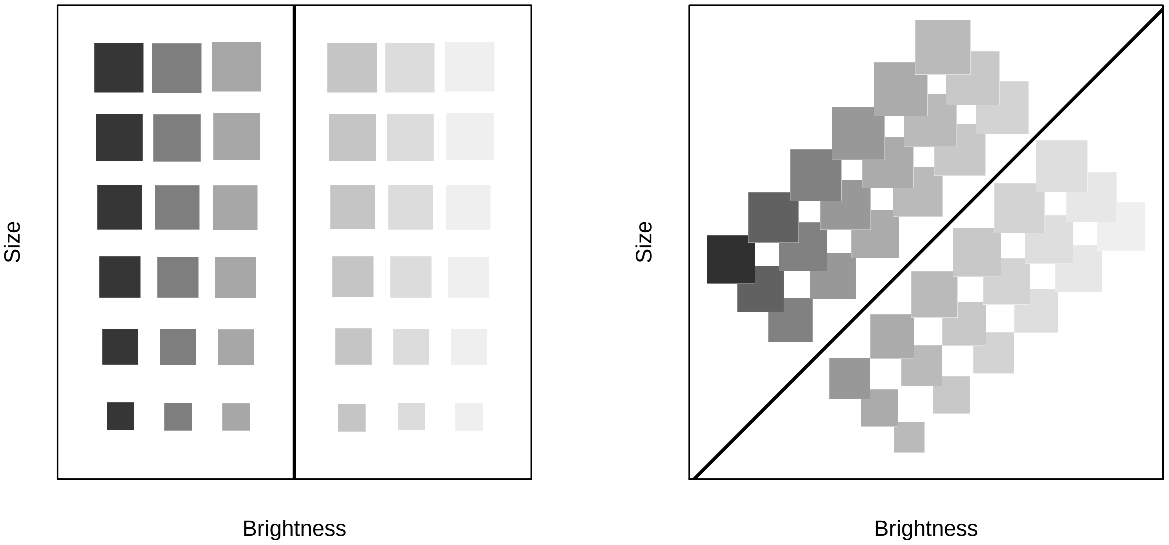

Figure 1: a) The unidimensional category structure with a rule based on size. b) The information-integration category structure with a positive category boundary.

tive to a well-matched rule-based category learning condition (Carpenter, Wills, Benattayallah, & Milton, in press). The medial temporal lobe has long been considered critical for explicit memory. Therefore, Carpenter et al.'s results suggest that information-integration category learning involves explicit memory processes to a greater extent than rule-based category learning.

One hypothesis that explains these two findings is that both rule-based and information-integration category structures are learned through the application of simple rules but that, for information-integration ones, those rules are supplemented by explicit memory of specific examples. For example, participants may store the examples that are exceptions to the simple rule. This hypothesis is also consistent with other evidence that found that participants who use a rule-plus-exception strategy have greater recognition memory for the exceptions to a simple category rule (Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004). As participants learning information-integration categories would have to remember more exceptions than those learning a rule-based structure, we would also expect greater recognition performance for information-integration learners.

However, this evidence is not conclusive. Neither Edmunds et al. (2015) nor Carpenter et al. (in press) directly measure explicit access to category knowledge. Further, some of the results of Carpenter et al. are at variance with a previous neuroimaging study of rule-based and information-integration category learning (e.g. Nomura et al., 2007). Carpenter et al. argue that the differences between their study and that of Nomura et al. are due to methodological problems with the Nomura et al. study. This argument is supported by the fact that Carpenter et al.'s results are broadly in line with the only other closely-related neuroimaging study (Milton & Pothos, 2011). Nevertheless, more direct evidence is needed.

In the following experiment, we directly examined recognition memory for exemplars in rule-based and information-integration category learning. Recognition memory is commonly assumed to be a test of explicit memory processes (Gabrieli & Fleischman, 1995). If rule-based structures are learned explicitly and information-integration category structures are learned implicitly, as predicted by COVIS, then one would predict, if anything, better recognition memory performance for participants in the rule-based condition than for participants information-integration condition.

In contrast, our hypothesis is that that participants learn information-integration category structures explicitly using simple rules bolstered by memory for exceptions to those rules. This strategy would allow participants in the information-integration condition to score as highly as if they used the optimal diagonal decision bound, however it would also increase demands on recognition memory as participants would have to remember the exceptions. An alternative similar hypothesis is that participants in the information-integration condition may be using complicated rule strategies such as a conjunction rule. If this is the case, participants would still have to pay more attention to stimulus features information-integration than in the unidimensional condition as they would have to compare each stimulus to multiple boundaries. Whereas, for the unidimensional condition they only have to focus on one stimulus dimension. Either way, we would expect to see better recognition memory in the information-integration condition than in the rule-based condition.

Method

Participants

Forty-two undergraduate psychology students were recruited from the Plymouth University participation pool and compensated with partial course credit.

Stimuli and category structures

The stimuli were 36 grey squares that varied in brightness and size displayed on a white background. The stimuli seen by each participant depended on which category structure they learned.

Half the participants were randomly assigned to learn a unidimensional rule-based category structure and the other half to learn an information-integration category structure as illustrated in Figure 1. The orientation of the category boundaries in abstract stimulus space were counterbalanced within conditions resulting in two unidimensional category structures–with a rule based solely on either the brightness (11 participants) or size of the square stimuli (10 participants)–and two information-integration category structures–where the optimum boundary had either a positive (10 participants) or negative gradient (11 participants). In addition, the stimuli were log-scaled so that all adjacent stimuli were approximately equally perceptually discriminable.

The abstract representation of the information-integration positive category structure is identical to that used by Spiering and Ashby (2008) with 6 stimuli added to bring the total number of stimuli up to 36. These stimuli were added to facilitate the random selection of a third of stimuli as "unseen" items for the recognition task. The remaining category structures are rotations ($\pi/4$, $\pi/2$, $3\pi/4$ rad) of this original structure around the origin and then translated so that 'center of gravity' of the points remained the same.

Procedure

The experiment was split into three phases: category training, recognition test and finally, category test.

Category training In this phase, participants were trained on two thirds of the available stimuli. The training stimuli were selected randomly for each participant subject to several constraints: 1) that those stimuli selected were symmetrical around the category boundary and 2) that no adjacent stimuli of similar difficulty were removed. In total there were 360 training trials, split into 3 blocks of 120 trials. In each block, the 24 stimuli were shown 5 times in a random order. On each trial, the participants looked at the stimulus until they made a response using either the "Z" key for Category A or the "/" key for Category B. Participants were unable to respond until at least 500ms had passed. Then, either "Correct" in green or "Incorrect!" in red was displayed for 500ms. A blank white screen was displayed between each trial for 500ms. Throughout the experiment, the labels "Category A" and "Category B" were displayed on the bottom left and right of the screen respectively. If participants took longer than 5000ms to respond, no corrective feedback was given, instead the message "PLEASE RESPOND FASTER" was displayed for 500ms.

Recognition test In this phase participants judged whether each stimulus was "old" and appeared in the training phase, by pressing the "O" key, or was "new" and had not been shown in the training phase, by pressing the "N" key. The words "New" and "Old" were presented on the bottom left and right of the screen respectively. After this, participants judged the confidence they had in their old-new judgement on a Likert scale that varied from 1 (=guessed) to 5 (=certain) by pressing the corresponding number key. Each of the 36 stimuli were presented three times in a randomised order. No feedback was given.

Category test In this phase, participants were asked to judge the category membership of all 36 stimuli, not just those they had seen in the category training phase. The procedure was identical to that of the training phase, apart from there was no feedback. Each of the 36 stimuli were presented three times in a random order.

Verbal report questionnaire At the end of the experiment, participants were asked to complete a questionnaire that asked them to describe in detail the strategy that they used. This was to determine whether the participants could explicitly report the strategy they used and whether any participants used a rule-plus-exception strategy. The questionnaire asked them to "Imagine that another participant was asked to complete the experiment as you did. What instructions would you give them so that they could exactly copy your pattern of responding?"

The verbal reports were coded by CERE and AJW.

Analysis

All data analyses were conducted in R (R Core Team, 2014). For every condition and phase of the experiment, all trials were removed for which the reaction times were outliers in that condition (i.e. outside 1.5 times the interquartile range above the upper quartile and below the lower quartile).

Results

One participant was removed from the unidimensional condition because their accuracy score was consistently below chance (i.e. 50%), resulting in 20 and 21 participants in the unidimensional and information-integration category structure conditions respectively.

Performance

Category learning We found a statistically significant difference in categorization accuracy at test, $F(1,39) = 13.51$, $p < .001$. Proportion correct was higher for the unidimensional category structure, $M_{UD} = 0.87$, $SD = 0.07$, than for the information-integration category structure, $M_{II} = 0.78$, $SD = 0.11$.

Recognition To determine overall memory performance, we calculated d' values for each participant.

We found that there was a statistically significant difference in d' between the two category structure conditions, $t(39) = 2.04$, $p = .048$. Specifically, d' was higher for the information-integration category structure, $d' = 0.01$, $SD = 0.02$, than for the rule-based category structure, $d' = 0.00$, $SD = 0.01$. Further, d' was significantly greater than chance in the information-integration category structure condition, $t(20) = 2.98$, $p = .007$, but not in the unidimensional category structure condition, $t(19) = 0.67$, $p = .511$.

Strategy analyses

The performance analyses above indicate a slight memory advantage for stimuli in the information-integration category structure condition compared to those in the unidimensional category structure condition. In this section, we investigate possible sources of this advantage.

Model-based analysis One possibility consistent with COVIS is that the category structure manipulation failed to result in a corresponding shift in category learning system. In other words, participants in the information-integration condition could have been using the sub-optimum, explicit system. If this were the case, then it would not be surprising that participants had explicit memory for category information. This is always a concern for experiments in the COVIS literature. The usual solution is to conduct a model-based analysis to determine which strategies participants are using to learn the structure. If the majority of participants in the information-integration condition are identified by the analysis as using the optimum diagonal strategy, then proponents of COVIS would conclude that those participants are using the implicit system.

In the model-based analysis typically used in the COVIS literature (Ashby & Gott, 1988), four types of model are fitted to the data from each participant. These model types are qualitatively different types of optimum decision boundaries that split the stimulus space into two, with "Category A" responses on one side and "Category B" responses on the other.

Table 1: Strategies identified in each condition according to the model-based analysis.

	Strategies			
Condition	UD	GLC	CJ	RND
UD	14	2	3	-
II	6	10	4	1

Category structures: UD=Unidimensional, II=Information-integration. Models: UDX=Unidimensional based on brightness, UDY=Unidimensional based on size, GLC=General linear classifier, CJ=Conjunction RND=Random (both types).

The *unidimensional* models assume that the stimuli are categorised on the basis of a single stimulus dimension. In this case, there are two possible unidimensional models: the stimuli can be split either on the basis of brightness or size. A unidimensional rule based on brightness would be "Place the light squares in Category A and the dark ones in Category B". This would be represented in stimulus space as a vertical or horizontal line. This model has two parameters: the value at which the boundary crosses the axis and perceptual noise.

The *conjunction* models assume that participants make a decision for each stimulus dimension and then combines them to determine category membership. A conjunction rule in this case might be "Place the light, small squares in Category A. Everything else is in Category B." This model can have up to three parameters: a decision criterion on each dimension and a noise parameter. Four versions of the conjunction model were included corresponding to each corner of the stimulus space.

The *general linear classifier* (GLC) models assume that the decision boundary between the categories can be described by a diagonal line between them. This is the optimum strategy for the information-integration condition. This model can have up to three parameters: the gradient, intercept and a noise parameter.

Two types of *random* models were also included: unbiased and biased. In the unbiased model, it is assumed that for every stimulus the participant is equally likely to pick either category. There are no parameters in this model. In the biased random model, it is assumed that for every stimulus the participant in likely to ascribe it to Category A with a certain probability (i.e. 30%). This model has one parameter, the proportion of Category A responses, and is a more general version of the unbiased random model, for which the parameter is equal to 50%.

The data from each participant was fitted to each of these models. The degree of fit was measured by the Bayesian Information Criterion (Schwarz, 1978). The results from this analysis, which was performed using the grt package in the R environment Matsuki (2014), are reported in Table 1.

Here we can see that our data meets the criterion commonly used in the COVIS literature: the majority of participants in the information-integration condition have been found to be using the optimum GLC (or diagonal) strategy. Researchers in the COVIS framework would normally conclude from this that people in the information-integration condition were using the optimum implicit system.

Verbal reports The model-based analysis indicates that our data meet the minimum requirements for a COVIS study: there was an obvious effect of strategy type depending on which category structure participants learned. However, as mentioned in the introduction, the results of this analysis have been found to depend on the models included in the analysis

Table 2: Strategies identified in the verbal report questionnaire

	Strategies							
	UD	seqUD	CJ	CJ2	RuleX	Implicit	Other	None
UD	15	1	1	0	0	1	0	2
II	4	4	3	5	1	0	1	4

Category structures: UD=Unidimensional, II=Information-integration. Strategies: UD=Unidimensional, seqUD=Sequential unidimensional, CJ=Conjunction, CJ2=Double conjunction, RuleX=Rule-plus-exception.

(Donkin et al., 2015). Additionally, our previous research found that participants identified as diagonal classifiers could also report using complex rule-based strategies (Edmunds et al., 2015). In this section we examine the strategies that participants report using to see if participants in the information-integration condition report using complex rule-based strategies.

Participants reported several different types of strategy. A report was classified as a *unidimensional* strategy if it described a rule based only of the stimulus dimensions such as "the dark squares are Category A and the light ones are Category B."

A report was classified as a *sequential unidimensional* strategy if the participants first categorised the stimuli at the extreme ends of one stimulus dimension and then defined a second unidimensional rule, on the other stimulus dimension, for the stimuli in the middle of the dimension. For example, a participant might say: "The very small stimuli were in Category A, and the very large in Category B. For the middle sized stimuli, the light ones were in Category A and the dark in Category B."

A report was classified as a *conjunction* strategy if the participant described an AND rule on the basis of two stimulus dimensions such as "Stimuli that are both small and dark are in Category A; else Category B."

A report was defined as a *double conjunction* strategy if the participant described two opposing corners of the stimulus space, but failed to define the other areas of the space. For example: "Large and dark patterns go into B. Small and light colours into A." As can be seen in this example, the participant fails to describe what category small dark stimuli would be in.

A report was classified as a *rule-plus-exception* strategy if the participant reported a simple rule with some exceptions. For example, "Light stimuli were usually Category A and dark stimuli Category B. However, one light medium sized stimulus was in Category B."

A report was classified as an *implicit* strategy if the participant recommended "not thinking too much" or to "rely on instinct" or similar phrases.

The inter-rater reliability for whether or not a participant reported a strategy was perfect. Similarly, both coders agreed perfectly on the strategies participants used.

As we can see in Table 2, no participants reported using an implicit strategy in the information-integration condition. This replicates our finding that the model-based analysis does not correspond well to the strategies participants report (Edmunds et al., 2015). Furthermore, only one participant in the information-integration condition used a rule-plus-exception strategy. This indicates that the advantage in memory may not be due to the use of a particular strategy, but because of the need in complex strategies to attend closely to the stimuli in order to categorise them by comparing to multiple decision bounds.

Discussion

Summary

A key dual-system model of category learning, COVIS, predicts that categorization is mediated by two competing learning systems: one explicit and one implicit. These two systems are hypothesised to optimally learn two different types of category structure. The explicit verbal system optimally learns rule-based category structures, whereas the implicit system optimally learns information-integration category structures. A key feature of this model is that category knowledge learned using the implicit system is unavailable to consciousness (Smith et al., 2015). In contrast, behavioral and neuroscientific work from out lab indicates that participants learning information-integration categories are aware of category knowledge and may be using explicit memory to facilitate categorization (Carpenter et al., in press; Edmunds et al., 2015). This experiment aimed to directly test these possibilities by comparing participants' performance on an old-new recognition task after learning either a unidimensional rule-based category structure, or an information-integration one.

Contrary to the predictions of COVIS, we found superior memory for exemplars after learning an information-integration category structure compared to a rule-based one. This indicates that participants may learn information-integration category structure using complex rule-based strategies rather than implicitly. This would result in superior exemplar memory for items when learning a information-integration category compared to a unidimensional structure as participants would have to attend more closely to the stimuli's features in order to compare them to multiple decision boundaries. Previous conclusions in the COVIS literature concerning the presence of implicit-like category learning may be due to an over-reliance on the assumption that a limited model-based analysis can provide evidence for implicit responding (Donkin et al., 2015).

One apparent limitation of the current study is that response accuracy in the category test phase is lower in the information-integration condition than in the rule-based condition. In an ideal comparison between rule-based and information-integration learning, the conditions would be matched for error rate. However, it seems likely that improving overall performance on the information-integration could only improve memory for the exemplars as this would involve using a more refined rule-based strategy. Thus, it seems unlikely that such a change would qualitatively alter our conclusions.

Another potential limitation is that recognition performance is poor in both conditions of the current experiment. This may be due to the stimuli being perceptually very similar to one another. Further work might increase discriminability by adding additional features that were not predictive to category membership. We are currently investigating this possibility.

In conclusion, this experiment finds preliminary evidence that participants learning the information-integration category structure do so explicitly. This conclusion is in contrast to the prediction of the COVIS model, which assumes that the information-integration structure is learned implicitly.

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*(3), 442–481.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*(1), 149–178.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 1–13). New York: Cambridge University Press.

Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. N. (in press). A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping*.

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*, 933–948.

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 . *The Quarterly Journal of Experimental Psychology*, *68*(2), 1203–1222.

Gabrieli, J. D. E., & Fleischman, D. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science*, *6*(2), 76–82.

Matsuki, K. (2014). grt: General recognition theory [Computer software manual].

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS. *European Journal of Neuroscience*, *34*(8), 1326–1336.

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). *Systems of Category Learning. Fact or Fantasy?* (Vol. 54). Elsevier Inc.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, a. D., Gitelman, D. R., Parrish, T. B., ... Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*(1), 37–43.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 548–568.

Price, A., Filoteo, J. V., & Maddox, W. T. (2009). Rule-based category learning in patients with Parkinson's disease. *Neuropsychologia*, *47*(5), 1213–1226.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*(4), 534–553.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Smith, J. D., Zakrzewski, A. C., Herberger, E. R., Boomer, J., Roeder, J. L., Ashby, F. G., & Church, B. A. (2015). The time course of explicit and implicit categorization. *Attention, Perception, & Psychophysics*, *77*(7), 2476–2490.

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science*, *19*(11), 1169–1177.

Stanton, R. D., & Nosofsky, R. M. (2007). Feedback interference and dissociations of classification: evidence against the multiple-learning-systems hypothesis. *Memory & Cognition*, *35*(7), 1747–1758.

Zaki, S. R., & Kleinschmidt, D. F. (2014). Procedural memory effects in categorization: evidence for multiple systems or task complexity? *Memory & Cognition*, *42*(3), 508–24.