Simulation of Explicit and Implicit Category Learning: Evidence for a Single Learning System

Frédéric Dandurand (fdandu@ego.psych.mcgill.ca)
Department of Psychology, McGill University,
1205 Dr.Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada

Abstract

Previous experiments (Waldron & Ashby, 2001) showed that category learning was differentially impaired by a concurrent Stroop task, depending on the type of rule used to categorize items. Learning was more impaired for simple explicit rules than for complex implicit rules. The present simulation suggests that the multiple learning systems hypothesized by Waldron and Ashby are not necessary to explain their results because a single learning system provides a parsimonious account of the data. In this model, the harder of two concurrent tasks determines learning time. Therefore, combined task complexity explains why the concurrent Stroop task impairs learning in the explicit group more than in the implicit group.

Introduction

Categorization is an important cognitive task (e.g., Harnad, 2005). It is currently controversial whether categorization is a single process, or if there are multiple systems involved for different kinds of categorization tasks.

Waldron and Ashby (2001) performed an experiment with human participants that seemed to support the hypothesis of multiple learning systems. Participants had to categorize four-dimensional items. Two variables were manipulated. First, the type of rule determined the difficulty of the categorization task. In the explicit rule condition, items could be categorized according to a single input dimension. In contrast, under the implicit rule condition, three of the four input dimensions had to be integrated in order to determine category membership. They used term *explicit* because participants can typically verbalize simple, one-dimensional rules. By contrast, participants generally cannot verbalize complex, multi-dimensional *implicit* rules, even when they perform well on categorization tasks using those rules.

Second, Waldron and Ashby manipulated processing load. In a concurrent condition, participants performed a numerical Stroop task while learning to categorize. In a control condition, participants only had to learn to categorize.

Participants were randomly assigned to either the control condition or to the concurrent condition. Each participant was presented with four categories to learn (two explicit and two implicit) in random order. The set of features relevant for determining category membership was selected randomly, and changed for each of the four categories to be learned.

Results (Waldron & Ashby, 2001) are reproduced in Figure 1. Their statistical analyses of those results showed:

1. A significant main effect of **condition** (i.e., processing load) – "… showing that the concurrent task group needed more training to learn the category structures than did the control group." (p. 171)
2. A significant main effect of **rule type** – "…showing that, over all other conditions, explicit rules required less training than did implicit rules." (p. 171)
3. A significant interaction between **rule type** and **condition** – "…showing that the concurrent task produced greater interference with explicit rules than with implicit rules." (p. 171)

In addition, Waldron and Ashby found a significant improvement in performance by the explicit concurrent task condition as the experiment progressed. The probable cause is a reduction of concurrent task interference, because the Stroop effect is known to diminish with training (Stroop, 1935; MacLeod, 1991). As a result, the critical Rule Type by Condition interaction found "Early in Session" disappears by the end of the experiment ("Late in Session"). No other differences between Figure 1A and 1B are statistically significant.

Waldron and Ashby (2001) claimed that the observed rule type by condition interaction supports the existence of multiple learning systems because the concurrent task interferes with explicit learning, but not with implicit learning. They further claim that the differential improvement in explicit learning during the experiment also supports multiple systems. Finally, Waldron and Ashby argued that, because the secondary (Stroop) task is commonly thought to reduce processing resources available to the primary (categorization) task, category learning should be more difficult when the Stroop task is concurrently performed. Furthermore, complex, implicit rules require more processing resources than simple, explicit rules. Thus, they conclude that in a single learning system, processing of implicit rules should always be impaired at least as much as that of simple rules when performed concurrently with the secondary task. Finally, they say that this prediction is contrary to what was observed: the learning of explicit rules was more impaired under concurrent task that the learning of implicit rules.

These kinds of arguments based on interactions are commonly taken as evidence for multiple learning systems. An important example of such interactions is *double dissociations*. However, some simulations suggest that multiple learning systems are not necessary to account for them. In fact, Kello, Sibley and Plaut (2005) found that a single connectionist system could model double dissociation phenomena.

Similarly, this paper challenges Waldron and Ashby's claim that the interaction they found is evidence for multiple learning systems. A simulation of their experiment was performed using Cascade-correlation neural networks (Fahlman & Lebiere, 1990). Results suggest that multiple category learning systems might not be necessary to account for their data.

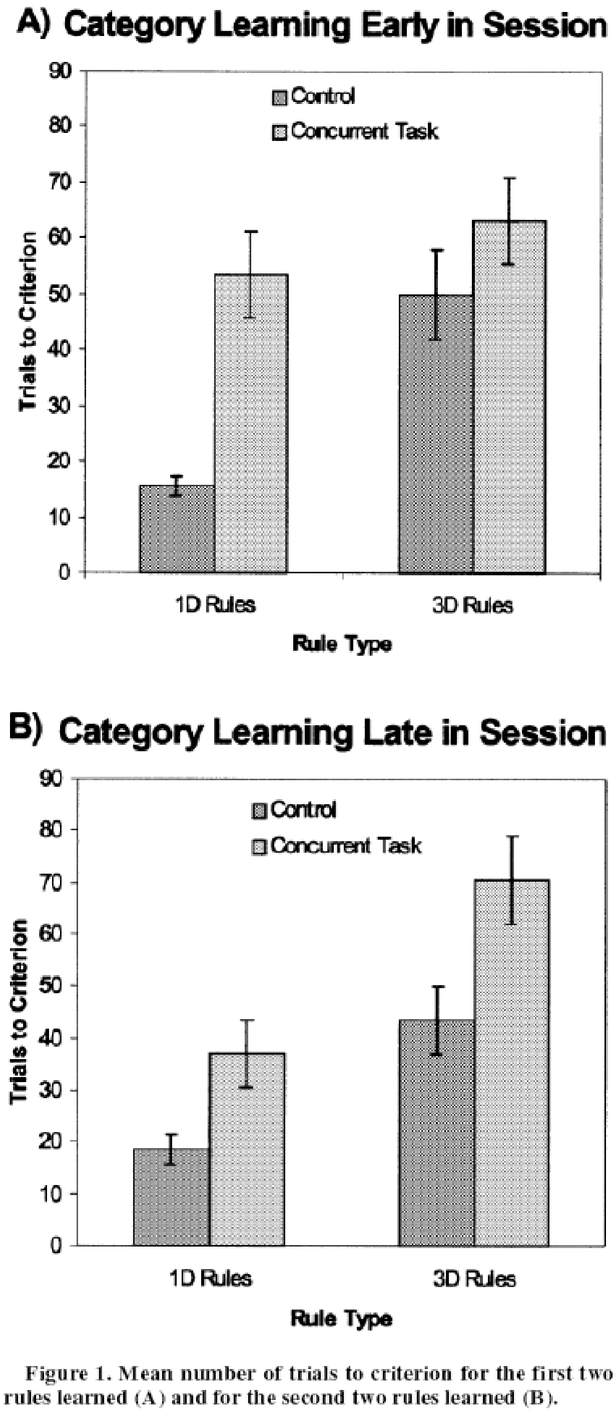

Figure 1 – Results from Waldron and Ashby (2001). Note that 1D rules were explicit, whereas 3D rules were implicit.

Note. From "The effects of concurrent task interference on category learning: Evidence for multiple category learning systems" by E. M. Waldron and F. G. Ashby, 2001, Psychonomic Bulletin & Review, 8 (1), p. 172. Copyright 2001 by The Psychonomic Society. Reprinted with permission.

Method

In the model, the categorization task has 4 inputs (binary features) and 1 output for a binary category decision (as in Waldron & Ashby, 2001). Like Waldron and Ashby (2001), two kinds of rules determined category membership: explicit rules where a single input feature determined the output (e.g., output is 1 if the second feature is 1), and implicit rules where 3 out of 4 features need to be integrated (output is 1 if any 2 of those 3 features are 1). The choice of input features used for categorization was random and different for each category. An example of implicit categorization task is presented in Table 2.

The concurrent task was a Stroop task similar to the one used by Waldron and Ashby (2001). There were four binary input units representing the numerical (N) and physical (P) size of each of two digits (D1 and D2), each coded as 0 for *small* and 1 for *large*. A fifth binary input unit coded an instruction (Instr.) to identify the larger digit based on its numerical value (0) or physical size (1). The output was binary coded to identify which digit was larger (0 for digit 1, or 1 for digit 2).

Table 1 – Stroop task patterns (N=16)

Inputs					Output
D1-N	D2-N	D1-P	D2-P	Instr.	
0	1	0	1	1	1
1	0	0	1	1	1
1	0	1	0	1	0
0	1	1	0	1	0
0	1	0	1	1	1
1	0	0	1	1	1
1	0	1	0	1	0
0	1	1	0	1	0
0	1	0	1	1	1
1	0	0	1	1	1
1	0	1	0	1	0
0	1	1	0	1	0
0	1	0	1	1	1
1	0	0	1	1	1
1	0	1	0	0	0
0	1	1	0	0	1

In the training set, 14 of the 16 patterns were questions about physical size, while 2 were about numerical size. Networks were presented with patterns in which the physical and numerical sizes were different, representing 4 different combinations: (Small and Large) x (Digit 1 and Digit 2), therefore some replications of patterns were necessary to match the 16 patterns of the categorization task.

Table 2 – Example of patterns for an implicit task (N=16). In this example, the 4th input feature is not used to determine category membership. Output is 1 if at least two out of the other three features are 1.

Inputs				Output
I1	I2	I3	I4	
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	0
0	1	1	0	1
0	1	1	1	1
1	0	0	0	0
1	0	0	1	0
1	0	1	0	1
1	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

For the simulations of the concurrent condition, the categorization and concurrent tasks were learned in parallel. As a straightforward way to model this, inputs and outputs of each task were concatenated, for a total of 9 and 2 respectively. 32 networks were trained in each condition for a total of 128 networks. Networks varied in the values of their initial weights, which were randomly selected. Networks trained in the early in session condition were reused to simulate the late in session condition, after connection weights were modified to partially reset the Stroop effect (see Initializing Networks section for details). The goal was to mimic as much as possible the methods used in Waldron and Ashby (2001).

In the simulations, the Stroop task was construed as a short-term "on the fly" learning to suppress default automatic response and to generate the unusual, but correct answer (evaluate differences in physical size). Because connection weights were initialized so that the network automatically generates the automatic response (evaluate differences in numerical size), the system had to unlearn this automatic response, and instead generate the correct response (evaluate differences in physical size). Stronger Stroop effects would manifest themselves in longer training time.

Similarly, the categorization rules learned were also short lived. The system had to constantly learn novel associations because new combinations of features determining category membership were randomly selected for each of the four categories.

Cascade-Correlation Algorithm

As mentioned, this simulation used the Cascade-correlation (Cascor) algorithm (Fahlman & Lebiere, 1990). Cascor is a general purpose neural network algorithm that successfully simulated a range of cognitive tasks including the balance scale task, acquisition of pronouns and learning of distance, time, and velocity concepts (Shultz, 2003). As opposed to standard backpropagation algorithms in which experimenters need to set network structure prior to training, Cascor is a constructive technique where network size expands as needed to solve a task.

Cascor networks begin with input and output units but no hidden unit. Training starts in the *output phase* during which Cascor minimizes the sum of squared error using some standard learning algorithm like QuickProp (Fahlman, 1988):

$$E = \sum_o \sum_p (V_{o,p} - T_{o,p})^2 \qquad (1)$$

where V is the activation of output o for pattern p, and T is the corresponding target value that the network is trying to learn.

If error reduction stagnates before the task is successfully learned, Cascor enters the *input phase*. In input phase, a set of candidate units compete for recruitment. Those units are typically sigmoids, and each one starts with different random input connection weights. By adjusting those weights, Cascor maximizes the covariance of each candidate units' outputs with the residual network error:

$$S = \sum_o \left| \sum_p (V_p - \overline{V})(E_{p,o} - \overline{E}_o) \right| \qquad (2)$$

where $E_{p,o}$ is the error at output unit o for pattern p, $\overline{E}_o$ is the average error at output unit o, V_p is the output unit activation for pattern p, and $\overline{V}$ is the average output unit activation.

When covariance maximization stagnates, the unit with the highest covariance is selected and connected to the output units, and thus becomes a new hidden unit[1]. Other units are discarded. Cascor then proceeds with another output phase with the newly recruited unit. Training alternates between output and input phases until target level of error is reached or training times out.

A major advantage of Cascade-Correlation over standard backpropagation is that determining network topology becomes part of the learning process, and is therefore automated. Although not necessarily optimal, network structures generated using Cascor tend to be relatively small (Fahlman & Lebiere, 1990).

[1] Note that Fahlman and Lebiere (1990) found that using covariance (S) worked better than using true correlation in most situations.

Initializing Networks

As noted, the Stroop effect diminishes with training (Stroop, 1935; MacLeod, 1991). To simulate this effect, a different initialization scheme for networks trained to simulate the early and late conditions was used.

Networks simulating the early phase were initialized so that numerical size outputs were given large connection weights between the numerical inputs and the output (called "numerical-bias weights"). Training thus involved inhibiting this automatic response (i.e., reducing the numerical-bias weights) and increasing weights between physical size inputs and the output.

By contrast, to simulate the late in session condition, network weights were reset to a value equal to the mean of early training weights and numerical-bias weights, as illustrated in figure 2. This reflected the fact that early training reduces the Stroop effect, but only partially.

To sum up, the experiment had two independent factors: (1) Condition (i.e., processing load) with two levels: Control (Categorization only) and Concurrent, and (2) Rule Type with two levels: Explicit and Implicit. Finally, there was one repeated factor, session with two levels: Early and Late.

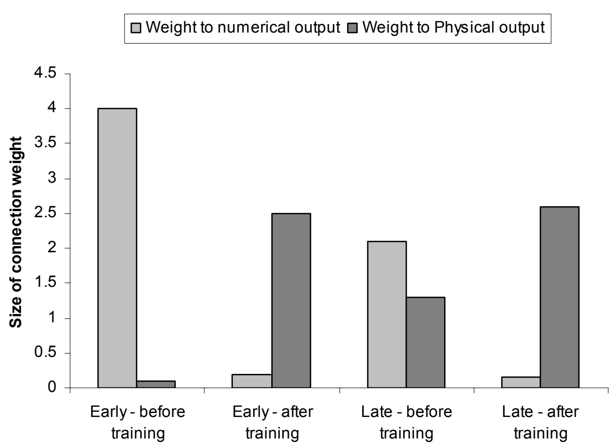

Figure 2 – Example of network initialization for the Stroop task. The first set of weights represents initialization of the network before training. A large weight to the numerical output (numerical-bias weight) yields the automatic (but incorrect) response. After training (second set of weights), the network learned to generate the correct answer, as shown by a large connection weight to physical output. In the third set, network weights are reset to the average of the previous two sets of weights (before and after training) to partially reset the automatic response. Finally the last set of weights (similar to the second one) indicates that the network is generating the correct response again after training late in session.

Results

Simulation results are presented in Figure 3.

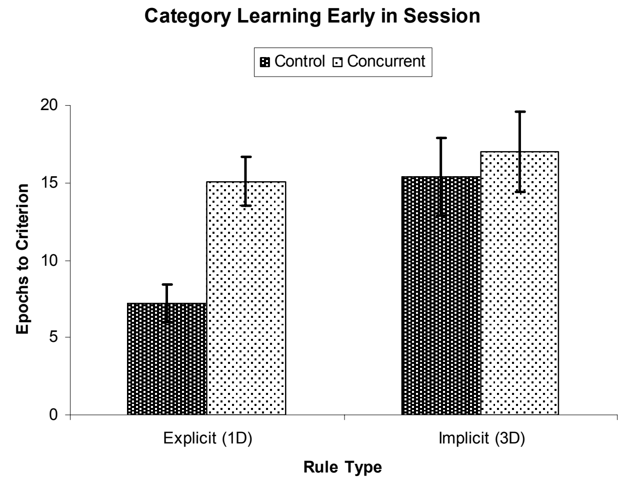

Figure 3A – Category Learning Early in Session. Error bars represent standard error.

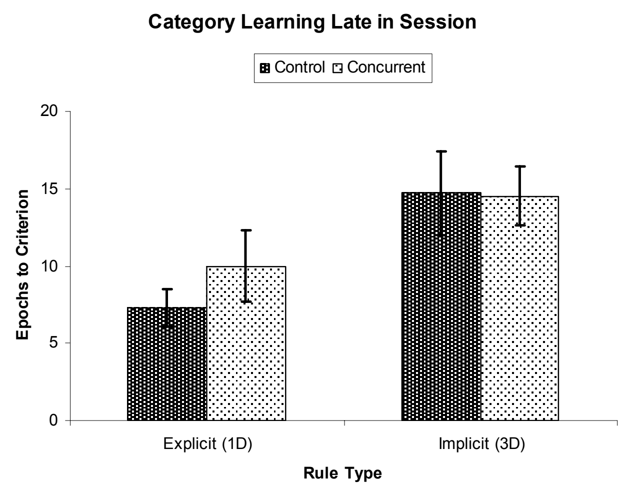

Figure 3B – Category Learning Late in Session. Error bars represent standard error.

Statistical Analyses

A mixed ANOVA was performed with Condition and Rule type as independent factors, and Session as a repeated factor. The following statistically significant effects were found:

1. Main effect of Condition (processing load): $F(1,124) = 129.8$, $p < 0.001$. Learning was faster in the control group than in the concurrent group. A main effect of Condition was also observed in human data.
2. Main effect of Rule Type: $F(1,234) = 437.8$, $p < 0.001$. Learning was faster for explicit rules than for implicit rules because explicit rules are simpler than implicit ones.
3. Condition by Rule Type interaction: $F(1,124) = 77.3$, $p < 0.001$. The model produced the same critical interaction that Waldron and Ashby (2001) found.

4. Main effect of Session: $F(1,124) = 67.4$, $p < 0.001$. Training was faster late in session than early, a logical consequence of reduction in the Stroop effect due to learning. This trend was present in Waldron and Ashby's work, although it was not significant.
5. Session by Condition interaction: $F(1,124) = 47.0$, $p < 0.001$. Again, as a result of the reduction of the Stroop effect, the concurrent group improved more than the control group during the late session.
6. Three-way interaction – Condition by Rule Type by Session: $F(1,124) = 10.8$, $p < 0.001$. The group that improved the most was the concurrent explicit condition, consistent with Ashby and Waldron (2001).

In the simulation, the only effect that was not significant was session by rule type interaction $F(1,124) = 2.9$, $p > 0.09$. Thus all significant effects in Waldron and Ashby's data were captured in the model, and it also produces two additional effects: a main effect of Session, and a Session by Condition interaction.

In short, when we compare Figures 1 and 3, we see that the pattern of simulation results is very similar to the one in Waldron and Ashby (2001) except that the F values were generally larger. This is a typical difference between human experiments and simulations because simulations have less error variance.

Discussion

The model was designed so that the categorization and concurrent tasks were learned in parallel. As a result, the number of epochs to criterion (success) was determined by the more complex task of the two.

Early in the session, the implicit and the Stroop task have about the same complexity. This explains why three of the four groups cluster around 15 to 17 epochs, while the system learns the easier explicit control task in about 7 epochs.

When the networks were retrained to simulate the late-in-session situation, categorization tasks remained as difficult as before because the content of the categories kept changing. However, the Stroop task was easier to learn because the weight initialization included a portion of the weights previously trained. Actually, by varying the proportion of trained vs. numerical-bias weights, the difficulty of the task can be varied from about zero (by taking 100% of trained weights, and 0% of numerical-bias weights) to as hard as at the beginning (by taking 0% of trained weights, and 100% of numerical-bias weights). Empirically, a 50% weighing resulted in a suitable level of difficulty.

In the simulation, changes in the difficulty of the Stroop task only affected the explicit concurrent condition. In fact, learning in the implicit concurrent condition was unchanged because the implicit task remained at the same difficulty level. This generally explains how the model captured differential performance impairment in rule learning in presence of the concurrent Stroop task, and the various statistical interactions observed.

Note that, from a computational modeling perspective, all these tasks are easy to learn using the Cascade Correlation learning algorithm. A single output phase was sufficient to learn the tasks, and thus no hidden units were recruited, indicating that these tasks are linearly separable. Because learning is so fast, differences of a few epochs can be relatively important. Consequently, results are sensitive to changes in parameter values such as inputs values, learning rates, and score thresholds.

In Cascade-Correlation networks, interference naturally occurs in recruited hidden units because all network inputs and previously installed units contribute to the weighted sum of input used to determine the level of activation of a given unit. As discussed above, although Cascade-Correlation can build complex network structures as it learns, no hidden units were necessary in this simulation to succeed at the tasks presented. As a result, the topology for this task is identical to a fully-connected backpropagation network with 4 inputs and 1 output.

I am currently working on a new model that captures interference effects using hidden unit recruitment. However, the current model does capture the limited capacity aspect of task concurrence by virtue of being built out of only 9 input and 2 output units.

In short, this single learning system accounts for the pattern of results in Waldron and Ashby (2001) because the harder task determines learning time. Combined task complexity (categorization + concurrency) explains why the Stroop task impairs the explicit task more than the implicit, and why the explicit concurrent group improves the most with training.

Concerns about Waldron and Ashby's study (2001)

Waldron and Ashby (2001) chose the numerical Stroop task because "Recent neuroimaging studies have shown that the anterior cingulate and dorsolateral prefrontal cortex are strongly activated in the Stroop (1935) task (Bench et al., 1993)" (p. 170). The same brain regions are active when an explicit rule is learned, but not when an implicit rule is learned.

Although this is a compelling reason for this choice, Waldron and Ashby (2001) did not control for task concurrence. Their hypothesis was that the Stroop task would interfere with the explicit, verbally-driven learning system causing more performance impairment to the explicit rule learning. However, perhaps concurrence by itself is sufficient to account for the pattern of data. Perhaps similar experiments should be performed involving other concurrent tasks varying in complexity and difficulty including some known not to activate the anterior cingulate and the dorsolateral prefrontal cortex.

Furthermore, Waldron and Ashby (2001) did not test the verbalizability of their rules. More specifically, after participants reached success criterion they should be asked what rule they are using to classify elements to verify the explicit/implicit nature of the rules.

Under a multiple learning system model, learning using the explicit learning system would be impaired with the concurrent task. Because learning of one-dimensional rules still occurs under Stroop task concurrence, perhaps the implicit learning system is responsible for such learning. Actually, the fact that performance level is very similar under impaired (concurrent) one-dimensional learning and three-dimensional rule learning is compatible with this claim. Under the multiple learning system model, participants should not be able to verbalize one-dimensional rules learned using the implicit system because the implicit system is not connected to language processing modules of the brain.

By contrast, an alternative explanation supporting a single category learning system is that differences in ability to verbalize rules are due to limitations in the system which is interpreting, extracting, or decoding what the single learning system has learned. It might be that, while an interpretive system is capable of extracting verbally-encoded rules for simple tasks such as one-dimensional rules, it can not do so for more complex tasks like the three-dimensional rule. This model therefore predicts that participants would be able to verbalize one-dimensional rules learned in a concurrent condition.

Other models of these data have been proposed, including COVIS (Waldron & Ashby, 2001) and Alcove (Nosofsky & Kruschke, 2002, see also Ashby & Ell, 2002 for a reply). COVIS posits different learning systems for explicit and implicit rules. ALCOVE requires setting four free parameters.

In short, the current model suggests that Waldron and Ashby's multiple learning systems are not necessary to cover the critical Rule Type by Condition interaction, that is, the fact that the concurrent Stroop task interferes more with the learning of explicit rules than implicit rules. A Cascade-Correlation model provides a simple and parsimonious account in a single learning system. This and other simulations (e.g., Kello et al., 2005; Nosofsky & Kruschke, 2002) suggest that we need to be careful about using interaction evidence to draw conclusions about complex cognitive systems.

Acknowledgments

This work was supported by the Lloyd Carr-Harris McGill Major Fellowship.

I would also like to thank Thomas R. Shultz and Kris Onishi for their input, guidance and support.

References

Ashby, F. G., & Ell, S. W. (2002). Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2002), *Psychonomic Bulletin & Review, 9* (1), 175-180.

Fahlman, S. E. (1988). Faster-learning variations on back-Propagation: An empirical study. *Proceedings of the 1988 Connectionist Models Summer School*, Morgan Kaufmann.

Fahlman, S. E., & Lebiere, C. (1990). The cascade correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 524-532). Los Altos, CA: Morgan Kaufmann.

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. In C. Lefebvre & H. Cohen (Eds.), *Handbook on Categorization*, Elsevier.

Kello, C. T., Sibley, D. E., & Plaut, D. C. (2005). Dissociations in performance on novel versus irregular items: Single-route demonstrations with input gain in localist and distributed models. *Cognitive Science, 29*, 627-654.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163-203.

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* (1), 169-174.

Shultz, T. R. (2003). Computational Developmental Psychology. Cambridge, MA: MIT Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 28,* 643-662.

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* (1), 168-176.