An investigation of the ordinal adequacy of COVIS, a dual-system model of categorisation.

Charlotte Edmunds

August 15, 2012

Abstract

This study examines a dual-system model of categorisation, COVIS (COmpetition between Verbal and Implicit Systems; Ashby, Paul, & Maddox, 2011), and it's ability to predict the results of three categorisation experiments: Waldron and Ashby (2001), Nosofsky, Gluck, Palmeri, McKinley, and Glauthier's (1994) replicateion of Shepard, Hovland, and Jenkins (1961) and Experiment 2 from Medin and Schaffer (1978).

COVIS consists of a verbal, rule-based system and a separate, implicit, procedural-learning based system that compete for control of responding. Although the assumptions of COVIS have been empirically tested, the ability of COVIS to predict experimental results has rarely been tested and these previous examples of model fitting have been somewhat flawed as they failed to use an optimisation algorithm to determine the value of COVIS's free parameters. This study aims to fill that gap.

We found that although COVIS was able to replicate the model fitting of Waldron and Ashby (2001) by Ashby et al. (2011), it did not do so using the implicit system, thereby contradicting the assumptions of COVIS. Similarly, the model fittings to Nosofsky et al. (1994) and Experiment 2 in Medin and Schaffer (1978) were unsuccessful and also failed to utilise the implicit system. The implications of this are discussed and future directions for investigation suggested.

The human race's ability to categorise objects by assigning them to different groups is not only highly adaptive (Ashby & Maddox, 2005) but is also integral to many other cognitive processes such as language development, perception and learning (E.g. Franklin & Davies, 2004; Tulving & Psotka, 1971). Many formal models of categorisation, those that explicitly detail a transformation, usually mathematical, from a set of independent variables to dependent variables (Wills & Pothos, 2012), have been developed in order to explain these experimental findings (?, ?). Although formal models can be time consuming and difficult to develop, they offer a huge advantage over other simpler theories: they allow precise predictions to be made and tested (?, ?). However, as there are more than one model, the problem then become how to determine their relative adequacy.

Wills and Pothos (2012) suggest evaluating them on the basis of the number of ordinal, irreversible successes in predicting experimental results that can be attributed to each model. They argue that models should be primarily assessed on whether they capture the ordinal pattern of the data set and only then should the quantitative difference between the model and data be minimised as measures of quantitative fit can often be misleading. For example, the models in Figure 1 both have the same SSE but most would agree that model 1B provides a better description of the data than 1A. It is also easy to overfit models if only using a quantitative measure of fit. Expand?

Wills and Pothos (2012) also argue that these successes need to be irreversible. An

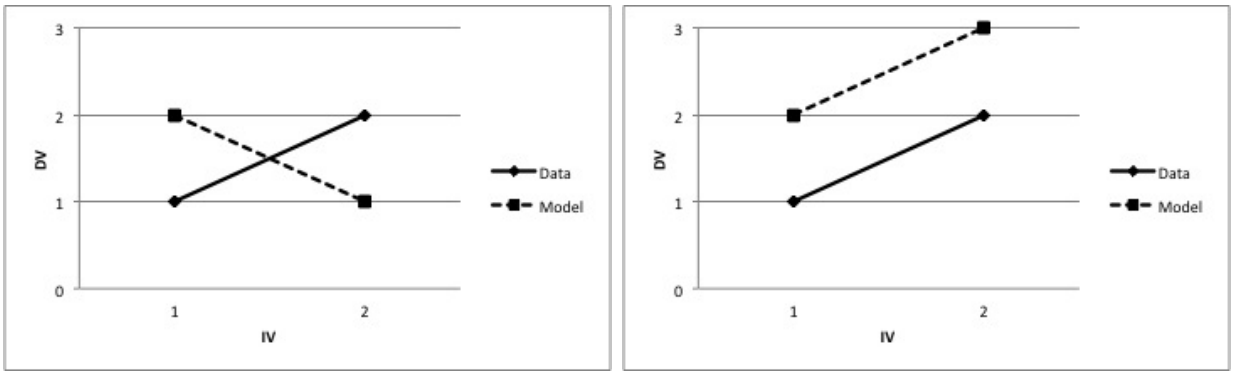

Figure 1. A. An ordinally incorrect model, SSE=2; B. An ordinally correct model, SSE=2. IV=Independent Variable, DV= Dependent Variable.

irreversible success is a successful prediction that doesn't require the use of arbitrarily variable parameters, those that are not held constant across all experimental results that the model claims to be able to describe. Arbitrary parameters cause problems, especially if by varying they cause ordinal changes in the output of the model. It may therefore appear that the model can accurately predict the data when in fact it can't.

This project aims to use these criteria to determine whether a particular model of categorisation, COVIS (COmpetition between Verbal and Implicit Systems Ashby et al., 2011), is capable of capturing the results from several experimental results central to the categorisation literature: Nosofsky et al.'s (1994) replication of Shepard et al. (1961) and Experiment 2, which uses the 5-4 category structure, from (Medin & Schaffer, 1978). COVIS has been chosen on the basis that it is an extremely popular model in the literature, however, the process of model fitting has only occasionally been applied and never to experimental data of normal, rather than patient, participants. The previous applications of COVIS to experimental data have all failed to use an optimisation technique, instead they have found reasonable guesses using a grid-search (E.g. Ashby et al., 2011; Helie, Paul, & Ashby, 2011; Hélie, Paul, & Ashby, 2012)

COVIS contains two systems of category learning that compete to control responding: a verbal, rule-based system and an implicit, procedural-learning system (Ashby et al., 2011). The verbal system is hypothesised to use explicit logical reasoning to test simple verbal rules, such as yellow stimuli are assigned to Category A whereas blue stimuli are assigned to Category B. On the other hand, the implicit system slowly adjusts the degree of connectedness from stimulus to a particular category and uses this in order to base a categorisation judgement. A competition system then takes information from both of these and, on the basis of the degree of trust in each system and the confidence of each system in their response, decides which system should control responding on that particular trial. This is formalised by COVIS in the following way.

The verbal system

The verbal system selects the most salient rule from the set of all possible verbal rules in order to determine a categorisation response. The set of all possible rules is denoted by $\mathbf{R} = \{R_1, R_2, \ldots, R_m\}$. $\mathbf{R}$ usually includes all one-dimensional rules as well as some more complex rules that are formed from the one-dimensional rules using Boolean algebra. However, to simplify the problem in the simulations below, as in other applications of COVIS (E.g. Ashby et al., 2011), only one-dimensional rules are included in $\mathbf{R}$.

To determine a response on trial n using rule R_i, COVIS firstly calculates a discriminant value, $h_E(\underline{x})$, which indicates whether the stimulus is greater or less than the category boundary on a particular dimension. Suppose that each stimulus varies on r dimensions, then each stimulus is denoted by $\underline{x} = (x_1, x_2, \ldots, x_r)$. It follows that, if R_i is a one-dimensional rule and the relevant stimulus dimension is i, $h_E(\underline{x})$ is determined by

$$h_E(\underline{x}) = x_i - C_i \tag{1}$$

where C_i is a constant representing the decision criterion, the boundary between one category and the next. The categorisation response is then determined by the following decision rule

$$\text{Respond A on trial } n \text{ if } h_E(\underline{x}) < \epsilon\text{; respond B if } h_E(\underline{x}) > \epsilon$$

where ϵ is a normally distributed random variable with mean 0 and variance σ_E^2. σ_E^2 is dependent on the participant's memory of the decision boundary and their perception of the stimulus, and tends to increase based on the increase in variability of these factors on each trial.

The rule to be used on any trial is chosen based on the success of the rule used on the previous trial. If the categorisation response was correct on trial n, then the rule used

on that trial will always be used on trial $n+1$. However, if the response is incorrect the next rule is selected from the set $\mathbf{R}$ on the basis of each rule's current weight. The weight of each rule is dependent on the participant's past experience of the rule, the reward history of the rule, the participant's tendency to perseverate and their tendency to select unusual rules.

Let $Z_k(n)$ denote the salience of rule R_k on trial n. Then, $Z_k(0)$ denotes the initial saliences of the rules in $\mathbf{R}$. The initial saliences of rules are high if they have been often used and low if they have not. In typical applications of COVIS this means that all one-dimensional rules tend to have equal, relatively high saliences whereas any conjunctive or disjunctive rules have much lower saliences. The salience of each rule is adjusted after each trial depending on whether rule R_i resulted in a correct response or not. If the response was correct on trial n then the salience of the rule used on that trial is adjusted by

$$Z_i(n+1) = Z_i(n) + \Delta_C \tag{2}$$

where Δ_C is a positive constant that represents the perceived reward associated with the correct answer. However, if the rule resulted in an incorrect response on trial n then the salience of that rule decreases by the rule

$$Z_i(n+1) = Z_i(n) - \Delta_E \tag{3}$$

where Δ_E is a positive constant that represents the perceived cost of an error on any trial. All the remaining rules in $\mathbf{R}$ keep their saliences from the previous trial.

The salience of each rule is then transformed to produce a weight, $Y_k(n)$, according to the following

(1) For the rule, R_i, that was active on trial n,

$$Y_i(n) = Z_i(n) + \gamma \tag{4}$$

where γ is a positive constant that represents the participants' tendency to perseverate on the active rule, regardless of the feedback given. COVIS assumes that executive attention is mediated by the head of the caudate nucleus and that γ is inversely related to the level of dopamine in the basal ganglia.

(2) For a rule choosen randomly from $\mathbf{R}$, R_j, it's weight is adjusted by

$$Y_j(n) = Z_j(n) + \mathbf{X} \tag{5}$$

where $\mathbf{X}$ is a randomly distributed variable that has a Poisson distribution with mean λ. $\mathbf{X}$ represents the participant's tendency to select novel rules on each trial; the larger λ is, the more likely they are to switch rules. COVIS assumes that rule selection is mediated by the frontal cortex and that λ is related to dopamine levels in the cortex.

(3) For the remaining rules

$$Y_k(n) = Z_k(n) \tag{6}$$

Finally, the rule to be used on the next trial is selected with probability

$$P_{n+1}(R_k) = \frac{Y_k(n)}{\sum_{s=1}^{m} Y_s(n)} \tag{7}$$

The implicit system

The implicit system of COVIS is hypothesised to facilitate learning of categories based on the integration of two or more stimulus dimensions. Typically, these types of category structures are difficult, or even impossible, to describe verbally. The implicit system closely follows the hypothesised neuroscience of procedural learning. Therefore, it consists of a representation of sensory information that leads tot a hidden layer representing the striatum, which in turn leads to a decision making process in the

prefrontal cortex as shown in Figure 2.

The sensory cortex is modelled by COVIS as an ordered array of up to 10,000 units. Each unit responds maximally to one particular stimulus and respond to a lesser extent to somewhat similar stimuli. The activation of each unit is calculated mathematically by a Gaussian function of the distance between the unit's preferred stimulus and the stimulus currently displayed, $d(K, \text{stimulus})$. So, the activation in sensory cortical unit K on trial n is given by

$$I_K(n) = e^{-\frac{d(K,\text{stimulus})^2}{\alpha}} \tag{8}$$

where α is a positive constant that scales the unit of measurement in stimulus space. The larger α is the more similar the stimuli are. However, in all the current simulations it is assumed that the stimuli are not confusable. Therefore, each stimulus is assumed, as in previous applications of COVIS (E.g. Ashby et al., 2011; Hélie et al., 2012), to only activate a single unit in the sensory cortex.

The activation of striatal unit $\mathcal{J}$ on trial n is determined by the weighted sum of

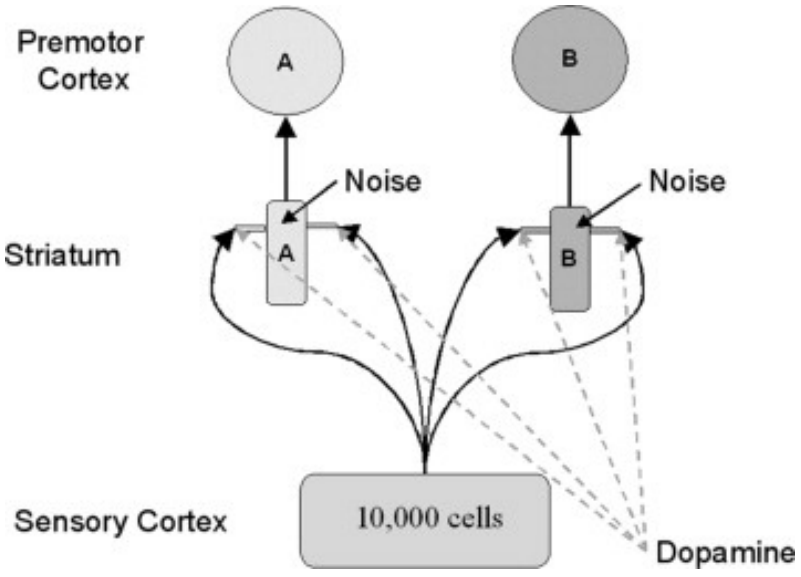

Figure 2. The architecture of the COVIS procedural system for a categorisation task with two categories. Source: Helie et al. (2011)

the activations of all sensory units that project to it, which is formalised as

$$S_{\mathcal{J}}(n) = \sum_{K} w_{K,\mathcal{J}}(n) I_K(n) + \epsilon \tag{9}$$

where $w_{K,\mathcal{J}}(n)$ is the strength of the synapse between cortical unit K and striatal cell $\mathcal{J}$ on trial n, $I_K(n)$ is the activation of sensory unit K on trial n and ϵ is normally distributed noise (with mean 0 and variance σ_P^2).

Then, the decision rule is

Respond A on trial n if $S_A(n) > S_B(n)$; otherwise respond B

The synapse strengths, $w_{K,\mathcal{J}}(n)$, are adjusted on each trial via reinforcement learning. The initial value, however, must be predetermined. In typical applications, $w_{K,\mathcal{J}}(0)$ is given by

$$w_{K,\mathcal{J}}(0) = 0.001 + 0.0025 \text{ U} \tag{10}$$

where U is a constant sampled randomly from a uniform $[0, 1]$ distribution. This means that all initial synaptic strengths will be between 0.001 and 0.0035 and will be randomly assigned. Then, $w_{K,\mathcal{J}}(n)$ is adjusted on each trial as follows

$$\begin{aligned} w_{K,\mathcal{J}}(n+1) = \; & w_{K,\mathcal{J}}(n) + \alpha_w I_K(n)[S_{\mathcal{J}}(n) - \theta_{\text{NMDA}}]^+[D(n) - D_{\text{base}}]^+[w_{\max} - w_{K,\mathcal{J}}(n)] \\ & -\beta_w I_K(n)[S_{\mathcal{J}}(n) - \theta_{\text{NMDA}}]^+[D_{\text{base}} - D(n)]^+ w_{K,\mathcal{J}}(n) \\ & -\gamma_w I_K(n)\{[\theta_{\text{NMDA}} - S_{\mathcal{J}}(n)]^+ - \theta_{\text{AMPA}}\}^+ w_{K,\mathcal{J}}(n) \end{aligned} \tag{11}$$

where if $g(n) > 0, [g(n)]^+ = g(n)$, otherwise $[g(n)]^+ = 0$.

It is not immediately clear what the terms of Equation 11 mean so the remainder of this section is spent deconstructing it. Broadly, the first line describes the conditions under which synapses would be strengthened whereas lines two and three describe the conditions

under which the connections would be weakened.

To start with α_w, β_w, γ_w, θ_{NMDA} and θ_{AMPA} are constants. The first three are learning rates whereas the constants θ_{NMDA} and θ_{AMPA} represent the activation thresholds for post-synaptic NMDA and AMPA glutamate receptors respectively, where $\theta_{\text{NMDA}} > \theta_{\text{AMPA}}$ because NMDA receptors have a higher threshold for activation than AMPA receptors.

Equation 11 requires that we specify the amount of dopamine released on every trial in response to feedback denoted $D(n)$. The amount of dopamine released is in turn dependent on the reward prediction error (RPE), which is determined by the following

$$\text{RPE} = \text{Obtained Reward} - \text{Predicted Reward} \tag{12}$$

The reward obtained on each trial is dependent on the feedback given. For example, for applications were all stimuli are rewarded or punished equally, obtained reward R_n on trial n is defined as $+1$ if correct feedback is given, 0 if no feedback is given or -1 if incorrect feedback is given. On the other hand, the predicted reward on each trial is calculated using a simplified version of the Rescorla-Wagner model (Rescorla & Wagner, 1972). Assuming that the participant has just responded for the nth time to some stimulus then the reward they should expect to receive is given by COVIS as

$$P_n = P_{n-1} + 0.025(R_{n-1} - P_{n-1}) \tag{13}$$

The dopamine release on each trial, $D(n)$, is then calculated from the RPE using the

following model

$$D(n) = \begin{cases} 1 & \text{if RPE} > 1 \\ 0.8\ \text{RPE} + 0.2 & \text{if } -0.25 < \text{RPE} \leq 1 \\ 0 & \text{if RPE } \leq -0.25 \end{cases} \tag{14}$$

The baseline dopamine level, D_{base}, which is the last remaining constant to appear in Equation 11, is 0.2 as can be seen from Equation 14 when the RPE = 0. This model was based on Bayer and Glimcher (2005).

Competition system

As only one response can be given on each trial, there needs to be a mechanism to resolve conflict between the explicit and procedural-learning systems. COVIS uses a combination of the confidence and trust it has in each system to decide which system there is to guide responding.

The confidence that each system has in its response is related to the degree of activation of each stimulus representation. The confidence in the explicit system equals the absolute value of the discriminant function, i.e. $|h_E(n)|$. If $|h_E(n)|$ is large then the stimulus is a long way from the decision bound and so the explicit system is more confident in its response, whereas if $|h_E(n)|$ is 0 then the stimulus is exactly on the boundary between two category and the explicit system has no confidence on its categorisation. The confidence in the procedural-learning system, $|h_P(n)|$, is equal to

$$|h_P(n)| = |S_A(n) - S_B(n)| \tag{15}$$

and follows a similar logic. If $|h_P(n)|$ is large then the procedural system favours one response much more than the other. However, if $|h_P(n)|$ is close to zero then both striatal units are activated equally and so the system has little confidence on its decision.

The degree of trust in each system is based on the past successes and failures of the system. The amount of trust in the explicit system is given by

$$\theta_E(n+1) = \theta_E(n) + \Delta_{OC}[1 - \theta_E(n)] \tag{16}$$

if the explicit system suggests a correct response on trial n. However, if the explicit system results in an incorrect response, the amount of trust in the explicit system on the next trial is given by

$$\theta_E(n+1) = \theta_E(n) - \Delta_{OE}\ \theta_E(n) \tag{17}$$

where Δ_{OE} is a parameter. The trust in the procedural system is given by

$$\theta_P(n+1) = 1 - \theta_E(n+1) \tag{18}$$

.

Then, COVIS emits the response suggested by the explicit system if

$$\theta_E(n)\ |h_E(n)| > \theta_P(n)\ |h_P(n)|$$

else it emits the response suggested by the procedural system.

Previous examples of model fitting

This example is the only time, to the author's knowledge, that COVIS has been tested on its ability to account for the categorisation behaviour of a purely normal sample. The other examples have tested COVIS's ability to account for experimental data that compares normal controls with patient samples, especially those patient samples which modulated dopamine levels (e.g.). This obviously plays to the strengths of COVIS as a neurally based model that can incorporate predictions about brain abnormalities. Were

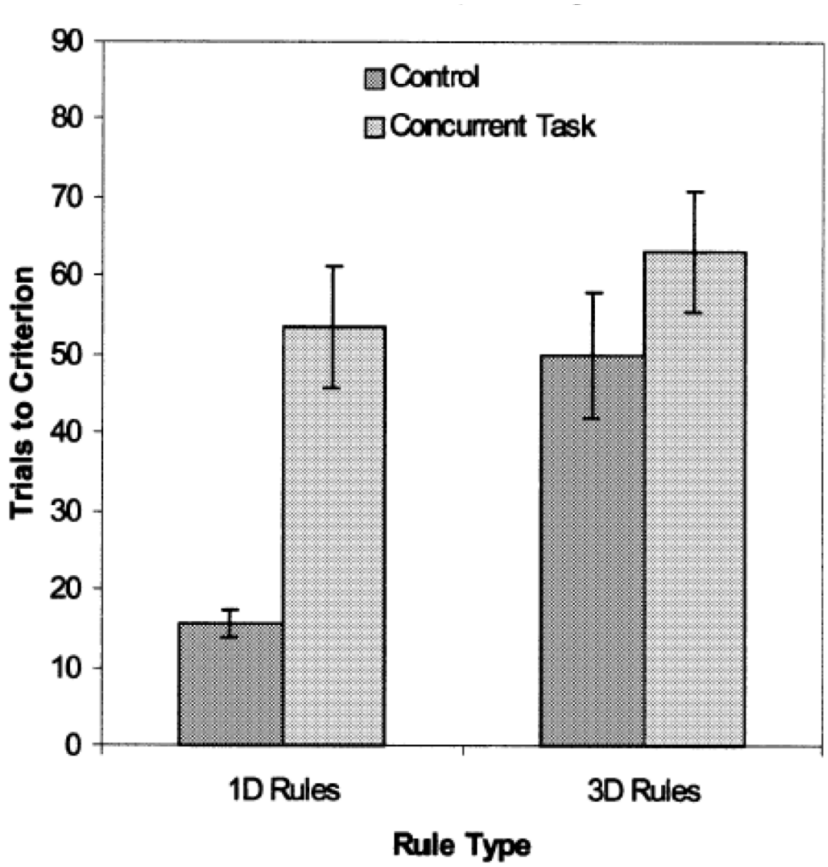

Figure 3. The results of Waldron and Ashby (2001). Source: Waldron and Ashby (2001)

they successful? However, it seems wise to first examine whether or not it can account for basic categorisation results in normal participants. Expand

Simulation 1: Ashby et al. (2011)

This simulation aims to replicate Ashby et al.'s (2011) use of COVIS to predict the results from Waldron and Ashby (2001).

Waldron and Ashby (2001) used a 2 (categorisation rule: rule-based, information-integration) by 2 (concurrent task: present, absent) within subject design. Sixteen stimuli were used that varied on four binary-valued dimensions (background colour, symbol colour, number of symbols and symbol shape). These were randomly presented

until either the participant made 8 correct responses in a row or reach 200 trials. The rule-based category structure used a category structure based on one stimulus dimension, for example, "Pick category A if the background colour is blue, otherwise pick category B". The information-integration category structure was determined as follows: one level of each stimulus dimension was assigned a numerical value of +1 and the other a value of -1. Then, one of the four stimulus dimensions was selected as irrelevant. Then the stimuli were assigned to one category if the sum of the values on the relevant dimensions was greater than zero and the other if it was less than zero. The categorisation task was either done on its own or with a concurrent Stroop-like task. In the concurrent task condition, two numbers were displayed before the stimulus to be categorised; one was physically larger and the other was numerically larger. After the participant had categorised the stimulus the participant was cued with either 'value' or 'size' at which point the participant had to respond with the appropriate number. On 85% of trials the physically larger number was numerically smaller. Ashby et al. (2011) were able to show that COVIS was able to predict results similar to those found by Waldron and Ashby.

However, Tharp and Pickering (2009) have criticised Waldron and Ashby's (2001) experiment, arguing that it is not a good test of information-integration category structures. Tharp and Pickering pointed out that a unidimensional rule would result in a 75% correct category assignment. This, along with a learning criterion that didn't check the participants' could correctly assign all 16 stimuli, meant that participants might be able to met the learning criterion easily without ever need to use an implicit learning system. They conducted a similar experiment, without the concurrent task, to ascertain whether the participants that successfully met the learning criterion in the information-integration condition were using an explicit, rule-based strategy and found that this was indeed the case in the majority of cases.

Therefore, this simulation aims to replicate Ashby et al.'s (2011) use of COVIS to predict Waldron and Ashby (2001) and also investigate which system provides the responses

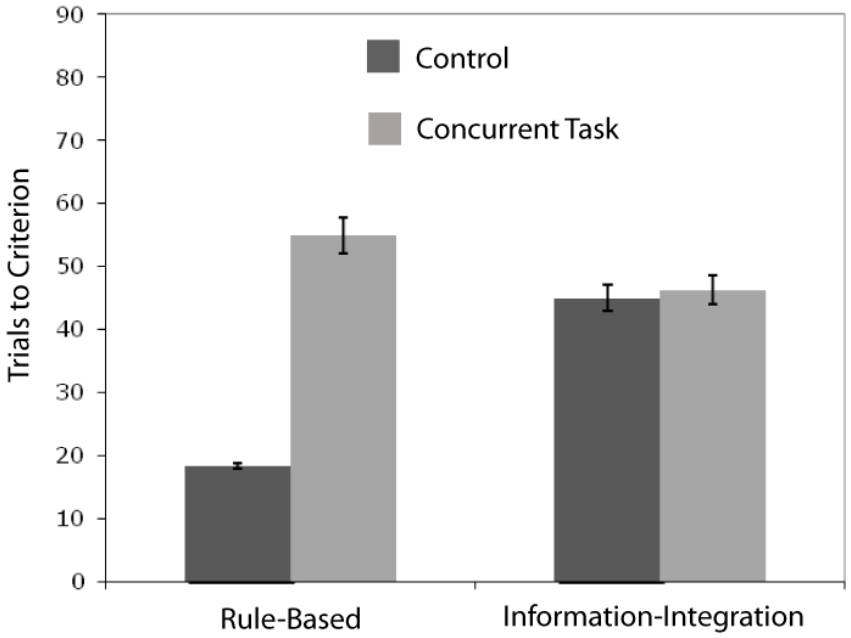

Figure 4. Ashby et al.'s (2011) fitting of the data from Waldron and Ashby (2001). Source: Ashby et al. (2011)

that result in a particular sequence reaching the learning criteria. COVIS would predict that in the rule-based category conditions, the verbal system would provide the optimum responses. On the other hand, in the information-integration category conditions, COVIS would predict that the implicit system would be more likely to provide the correct answers.

Replicating Ashby et al.'s (2011) fit to Waldron and Ashby (2001)

To obtain the values predicted by COVIS, one thousand stimulus sequences were randomly generated and the number of trials to criterion for each sequence were calculated by COVIS. Any stimulus sequences that failed to reach criterion were discarded. The number of trials to criterion of each stimulus sequence were then averaged across all stimulus sequences to determine COVIS's prediction for that experimental condition. No optimisation techniques were used to determine the values of the free parameters. Instead the values used by Ashby et al. (2011) were used (see Table 1).

Table 1. COVIS parameter values used in the simulations of Ashby et al. (2011)

Component	Parameter	Control	Dual-task
Explicit system	Δ_C	0.0025	Same
	Δ_E	0.02	Same
	γ	1.00	20.00
	λ	5.00	0.50
	$Z_k(0)$, for $k = 1, \ldots, 4$	0.25	Same
Procedural system	D_{base}	0.20	Same
	α_w	0.65	Same
	β_w	0.19	Same
	γ_w	0.02	Same
	θ_{NMDA}	0.0022	Same
	θ_{AMPA}	0.01	Same
	$w_{\max}$	1.00	Same
	σ_{P}	0.0125	Same
Competition	Δ_{OC}	0.01	Same
	Δ_{OE}	0.04	Same

Model Fitting Results

COVIS predicted results similar to those found by Ashby et al. (2011). However, as can be seen in Figure 5, COVIS failed to capture the ordinal properties of the data set.

There was also no significant difference between the rule-based and information-integration category conditions with respect to the proportion of stimulus sequences that met the learning criteria whilst using the implicit system ($\chi^2(8, 3220) = 11.14, p = .194$). This directly contradicts the assumptions of COVIS and supports the findings of Tharp and Pickering (2009).

Discussion

The results of this replication have indicated two flaws in COVIS's prediction of Waldron and Ashby (2001). The first is that this replication is unable to capture the ordinal properties of Waldron and Ashby's data. As COVIS is being primarily assessed on whether it can capture the ordinal properties of experimental results (as per Wills &

Pothos, 2012), this would indicate that the Waldron and Ashby experiment is outside COVIS's explanatory scope. However, it is possible that the choices of parameters were incorrect and that repeating the model fitting using an explicit optimisation process may result in COVIS making the correct ordinal prediction.

Unfortunately, it is impossible to attempt a more rigorous model fitting to Waldron and Ashby (2001) due to the paucity of data in the original paper. Therefore, in order to determine whether COVIS can capture the ordinal predictions of similar categorisation types, the next simulation will attempt to model Nosofsky et al.'s (1994) replication of Shepard et al. (1961) as the categorisation types I and IV in these studies are very similar to the ID and II categories used in Waldron and Ashby. Unfortunately, this experiment doesn't include a dual task condition so although successfully predicting this data would add support to the argument COVIS is capable of predicting Waldron and Ashby's

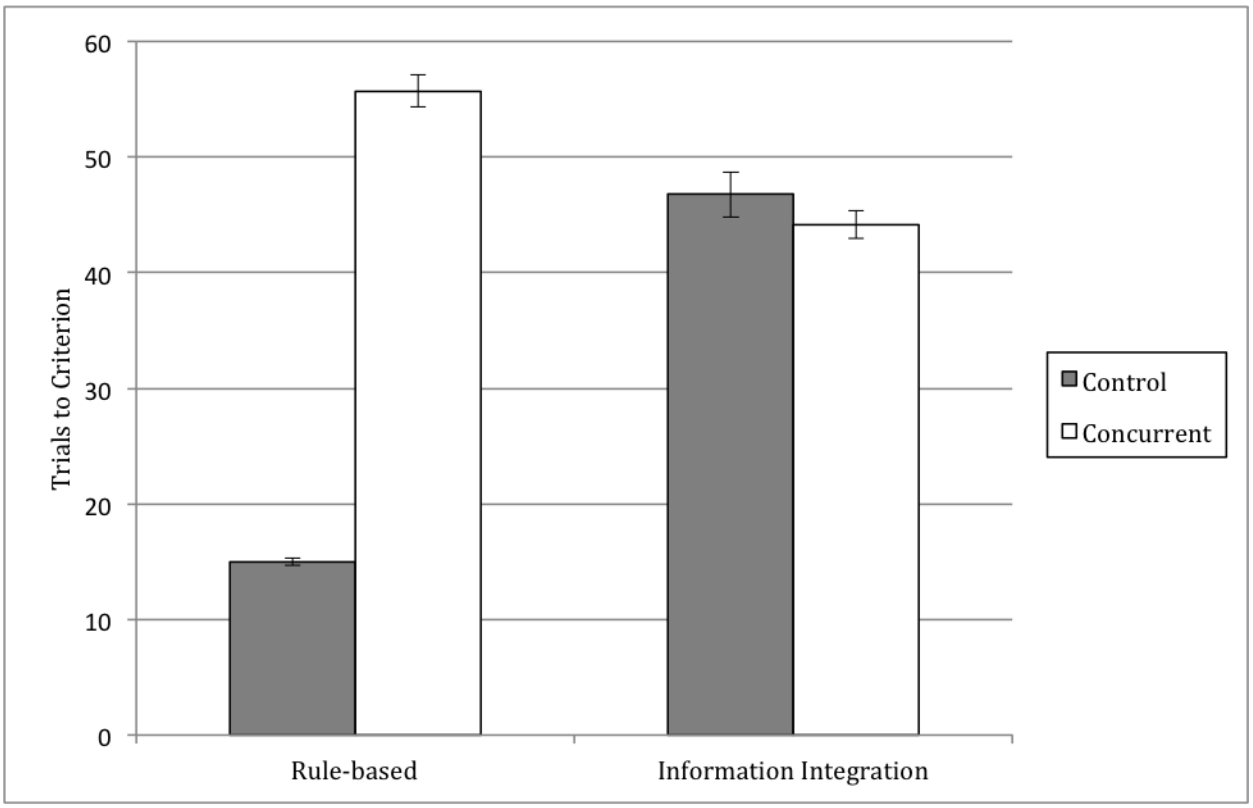

Figure 5. A replication of Ashby et al.'s (2011) fitting of the data from Waldron and Ashby (2001)

experimental, it would by no means prove it.

The second problem in this replication is that COVIS used the verbal system in order to determine the correct categorisation responses in both the rule-based and information-integration conditions. This supports Tharp and Pickering's (2009) argument that the Waldron and Ashby (2001) experimental design is incapable of differentiating between the different strategies used by participants in order to determine the correct category. By modelling Nosofsky et al.'s (1994) replication of Shepard et al. (1961) there will be a better test of the implicit system. For example, using a one dimensional rule in the Type VI category structure would be correct no better than chance. Also, the learning criterion in Nosofsky et al. tests the participants' knowledge for the category assigned to all eight stimuli.

Simulation 2: Nosofsky et al. (1994)

This simulations aims to test COVIS on its ability to predict the learning curves from Nosofsky et al.'s (1994) replication of Shepard et al.'s (1961) experiment. As well containing categorisation types similar to those used by Waldron and Ashby (2001), these studies are central in the categorisation literature and are commonly used by modern researchers as a standard by which categorisation models can be assessed (see for example Kruschke, 1992; Love, Medin, & Gureckis, 2004; Nosofsky & Palmeri, 1996). It includes a broad sample of possible category structures ranging from the very easy (Type I) to the more difficult (Type VI). Therefore, the experimental data from these studies provides an excellent test of COVIS's ability to mimic a wide range of categorisation behaviours. Nosofsky et al.'s (1994) replication was chosen as the data was more robust as it came from more trials and more participants.

In these studies, participants had to learn to assign stimuli to one of two categories using feedback. The stimuli varied on three binary-valued dimensions: shape (triangle or

Table 2. The six categorisations used in Shepard, Hovland and Jenkings (1961) and Nosofsky et al. (1994)

Stimulus	Type I	Type II	Type III	Type IV	Type V	Type VI
0 0 0	A	A	A	A	A	A
0 0 1	A	A	A	A	A	B
0 1 0	A	B	A	A	A	B
0 1 1	A	B	B	B	B	A
1 0 0	B	B	B	A	B	B
1 0 1	B	B	A	B	B	A
1 1 0	B	A	B	B	B	A
1 1 1	B	A	B	B	A	B

square), type of interior lines (solid or dotted) and size (large or small) resulting in eight possible stimuli. These stimuli were then categorised according to 6 different rules as shown in Table 2. The learning criteria was four consecutive sub-blocks of eight trials correct or until 400 trials had been reached.

The major finding of both Nosofsky et al. (1994) and Shepard et al. (1961) were that Type I categorisations were learnt the fastest, followed by Type II, then Types III, IV and V with the most difficult categorisation to learn being Type VI. This resulted in the learning curves displayed in Figure 6. Therefore, this simulation primarily aims to replicate this order of difficulty, with a secondary aim to demonstrate that COVIS can also predict results that are quantitatively close to it.

Following the demonstration above that COVIS used the verbal system to categorise the stimuli when modelling Waldron and Ashby (2001), it would also be interesting to determine which system COVIS uses to categorise each type. As stated by Waldron and Ashby (2001), the rule-based and information-integration category structures used in their experiment are similar to Types I and IV respectively. Therefore, they argue COVIS should use the verbal system in the Type I category structure whereas the implicit system should find the optimum strategy for the Type IV categories. In fact, as only one-dimensional rules are included in this version of COVIS, it would be reasonable to

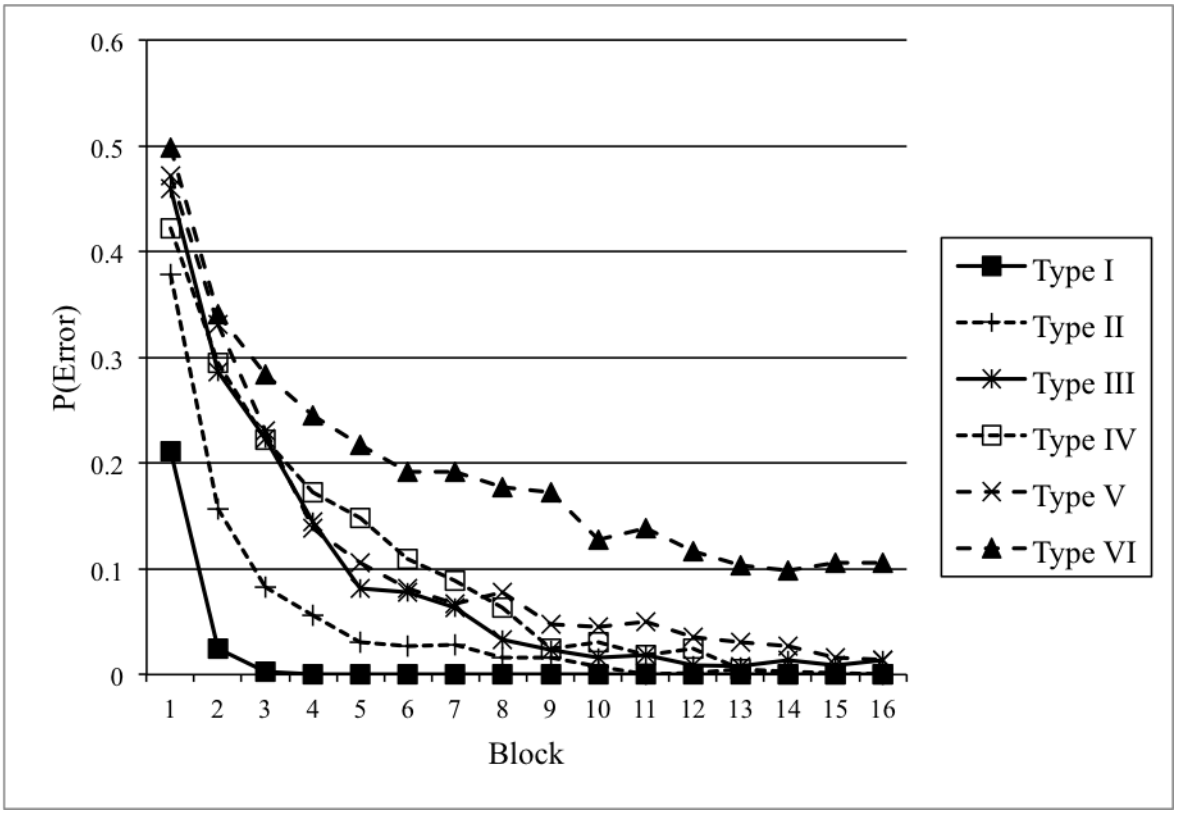

Figure 6. A graph of Nosofsky et al.'s (1994) replication of Shepard et al.'s (1961) experiment showing the average probabilities of errors for each categorisation type in each block of 16 trials

Table 3. COVIS parameter values used in the simulation of Nosofsky et al. (1994)

System	Parameter	Value
Explicit system	Δ_C	0.0051
	Δ_E	0.0293
	γ	3.1700
	λ	1.3204
	$Z_k(0)$, for $k = 1, \ldots, 3$	0.33
Procedural system	D_{base}	0.20
	α_w	0.0032
	β_w	0.4862
	γ_w	0.0206
	θ_{NMDA}	0.0130
	θ_{AMPA}	0.0022
	σ_{P}	0.7849
Competition	Δ_{OC}	0.0102
	Δ_{OE}	0.0451

predict that only Type I could be optimised using the verbal system. All the other rules would only reach the learning criteria if the implicit system controlled responding.

Fitting COVIS to the Learning Data from Nosofsky et al. (1994)

COVIS was only fitted to the first 16 blocks (256 trials) as after this point there minimal errors. The curves were fitted holding all parameters fixed across all six types of categorisation using an interior-point algorithm to minimise the sse between the predicted values and original data. Initial parameter estimates were varied to guard against local minimums and all parameter values were restricted to be positive and all, except γ and λ, were restricted to be less than 1. The parameters determined by the optimisation process are detailed in Table 3.

Model Fitting Results

Although, COVIS correctly predicted that the Type I category structure would be easiest to learn, it failed to predict any of the other ordinal properties of the learning

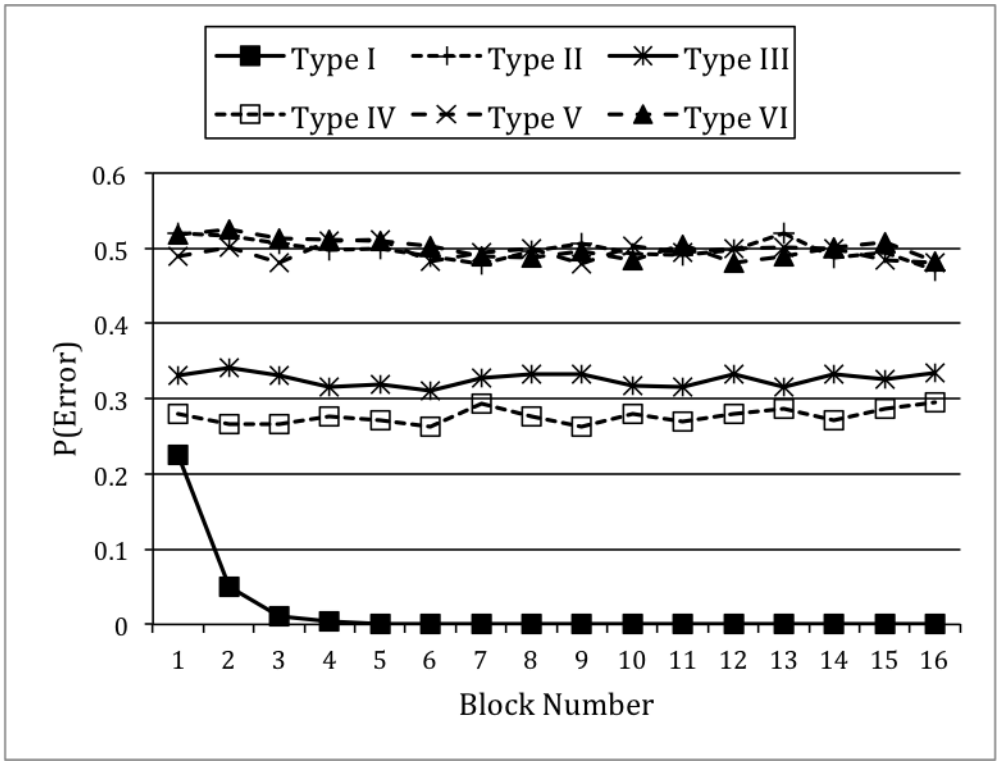

Figure 7. A graph showing COVIS's predicted values for Nosofsky et al. (1994) showing the average proportion of errors, for each category type, in each block of 16 trials

.

curves found by both Nosofsky et al. (1994) and Shepard et al. (1961), as can be seen in Figure 7. Unlike the original data set, COVIS predicted that Type IV would be the next easiest to learn, followed by Type III, and finally followed by Types II, V and VI together.

COVIS was able to replicate the learning curve of the Type I category (Figure 8). However, it was unable to learn to categorise the other, more complicated, category structures, as can be seen in Figures 9-13.

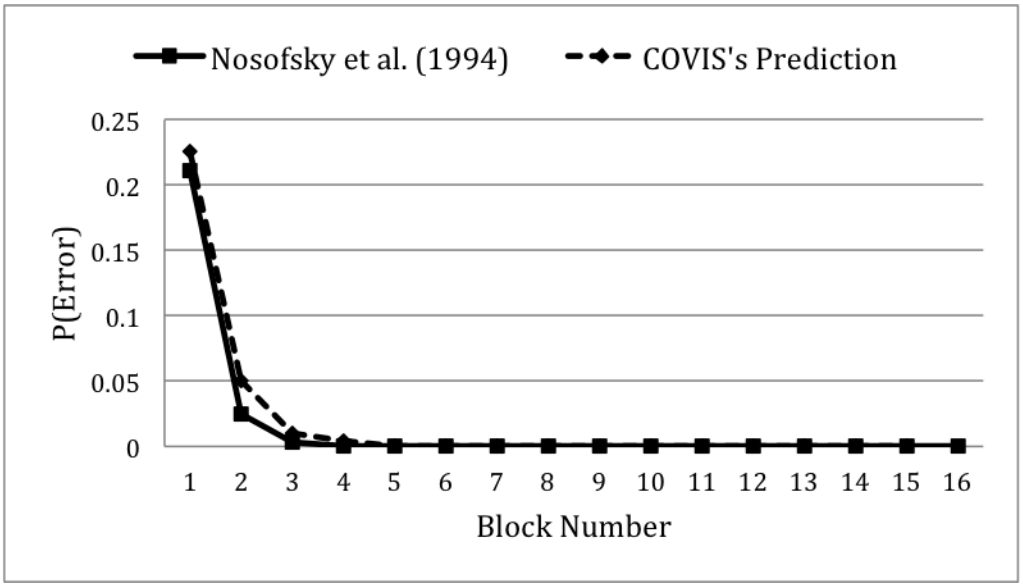

Figure 8. A graph comparing the learning curve for the Type I category structure predicted by COVIS with the experimental data from Nosofsky et al. (1994). Each point represents the average probability of error for each block of 16 trials.

Discussion

Although COVIS was able to capture the learning curve for the Type I category structure, it was unable to predict the learning curves for the other 5 category structures. In fact the predicted learning curves are consistent with COVIS solely using one-dimensional rules in the explicit system to determine category assignments. Table 4 shows the average error expected if the participants used one dimensional rules based on the first, second or third stimulus dimension. These correspond almost precisely with the learning curve predictions COVIS made. The obvious exception is Type I category

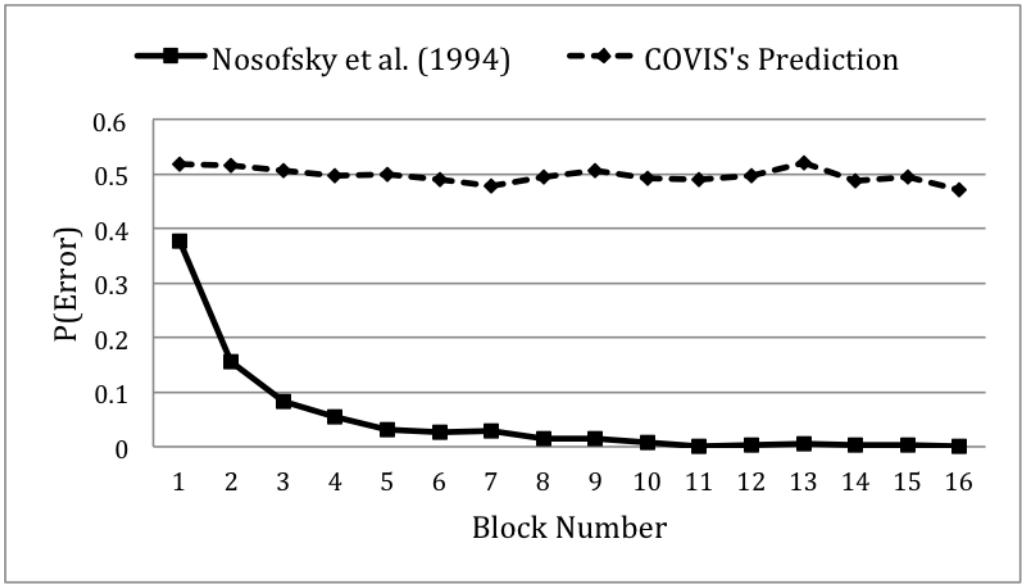

Figure 9. A graph comparing the learning curve for the Type II category structure predicted by COVIS with the experimental data from Nosofsky et al. (1994). Each point represents the average probability of error for each block of 16 trials.

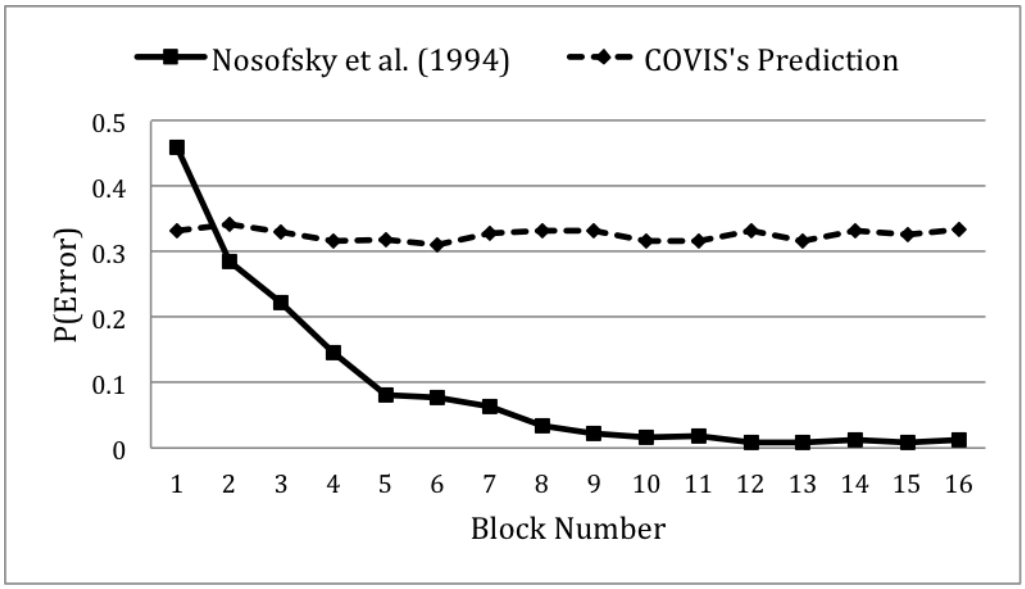

Figure 10. A graph comparing the learning curve for the Type III category structure predicted by COVIS with the experimental data from Nosofsky et al. (1994). Each point represents the average probability of error for each block of 16 trials.

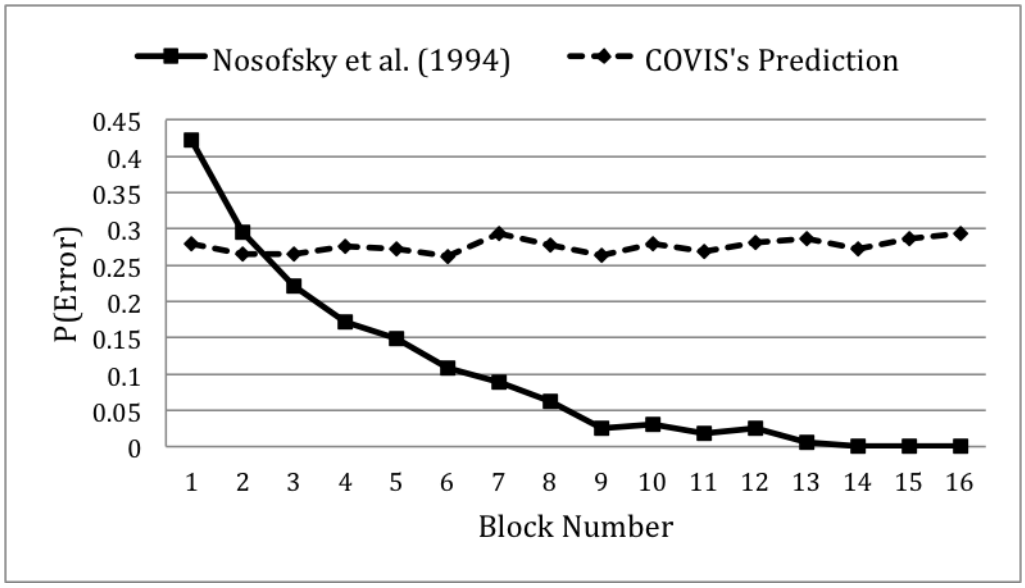

Figure 11. A graph comparing the learning curve for the Type IV category structure predicted by COVIS with the experimental data from Nosofsky et al. (1994). Each point represents the average probability of error for each block of 16 trials.

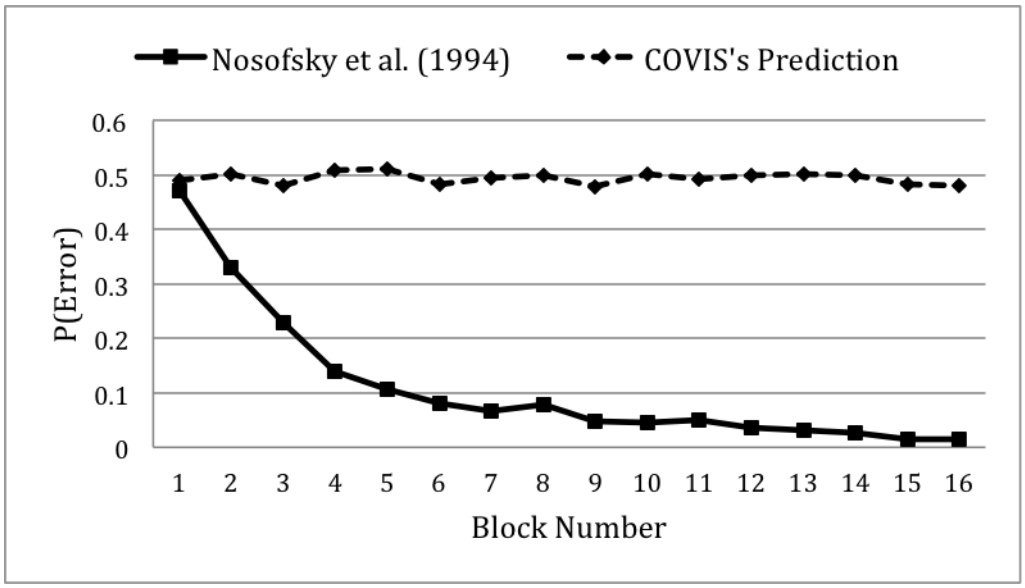

Figure 12. A graph comparing the learning curve for the Type V category structure predicted by COVIS with the experimental data from Nosofsky et al. (1994). Each point represents the average probability of error for each block of 16 trials.

Table 4. A table showing the fraction of errors you would expect if COVIS used a ID categorisation rule based on the 1st, 2nd or 3rd stimulus dimension

Rule based on:	Type I	Type II	Type III	Type IV	Type V	TypeVI
1st Dimension	0	0.5	0.25	0.25	0.25	0.5
2nd Dimension	0.5	0.5	0.25	0.25	0.5	0.5
3rd Dimension	0.5	0.5	0.5	0.25	0.75	0.5
Average Error	0.33	0.5	0.33	0.25	0.5	0.5

structure where COVIS was swiftly able to learn the one dimensional rule that provided the correct categorisation.

It should also be noted that COVIS failed to learn the optimum categorisation rule, when available, for any category structure apart from Type I. The error curves that COVIS predicted for category types III and V show that COVIS failed to learn which was the optimum one-dimensional rule. For example, a one-dimensional rule based on the first stimulus dimension would result in 25% errors for Types III and V, rather than 33% and 50% respectively. This seems to indicate that the verbal system in COVIS is incapable of learning which one-dimnesional rule minimises errors. It can only learn the optimum rule if it completely eliminates errors such as in Type I.

So why didn't COVIS use the implicit system to guide responding? One possibility is that the implicit system in COVIS takes much longer to learn than the participants in real life and so hadn't learnt the correct associations by the end of the experiment. This seems unlikely due to the shape of the learning curves in Nosofsky et al. (1994); there is a steep drop in errors between blocks one and two which is unlikely to be due to a slow learning system. However, to check this possibility I increased the numbers of trials available for COVIS to learn in. The results can be seen in Figure 14, COVIS still failed demonstrate learning in Types II-VI.

Another possibility is that the explicit system should be less prominent in the model and therefore that the value for θ_E is incorrect. This is extremely unlikely as numerous

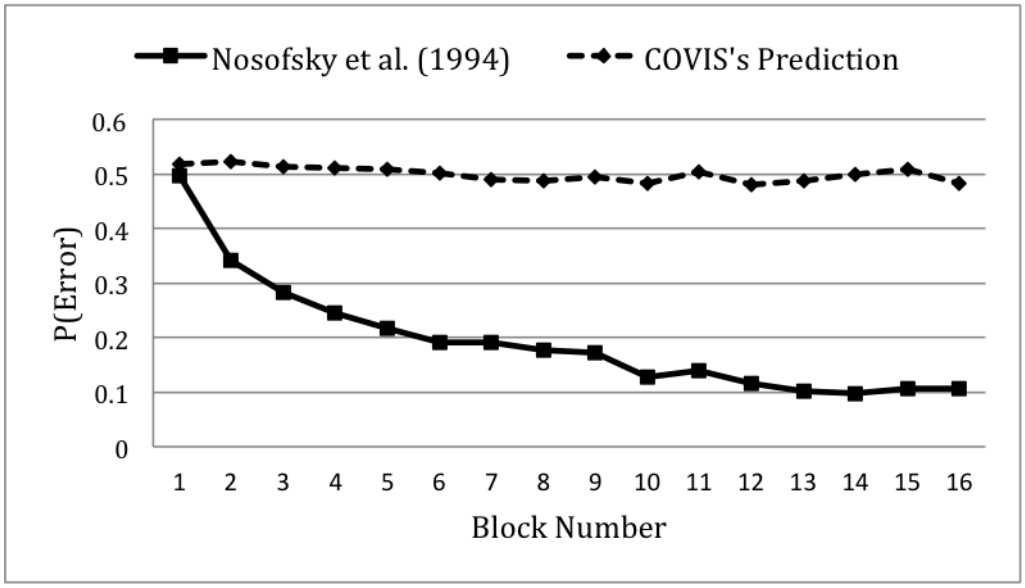

Figure 13. A graph comparing the learning curve for the Type VI category structure predicted by COVIS with the experimental data from Nosofsky et al. (1994). Each point represents the average probability of error for each block of 16 trials.

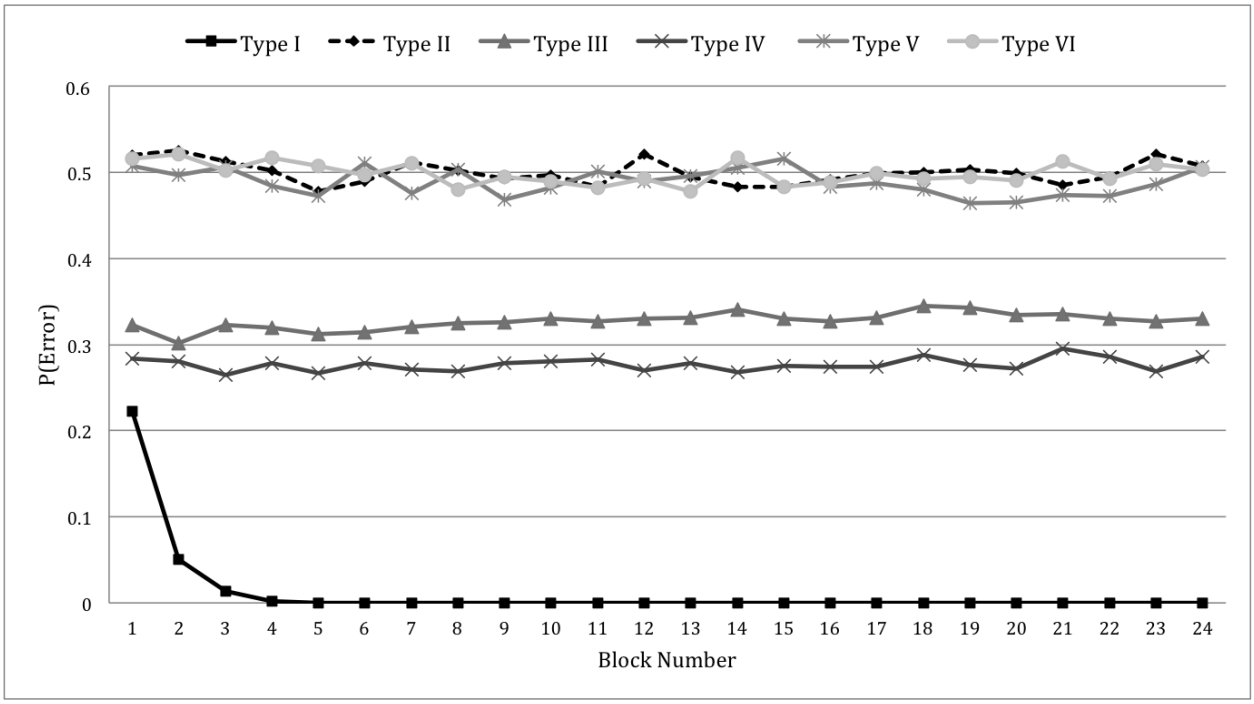

Figure 14. A graph showing all six learning curves predicted by COVIS over 24 blocks of 16 trials. Each point represents the average probability of error for each block of 16 trials.

experiments have demonstrated that participants start learning using the verbaliseable rules. However, when rerun with lower values of θ_E COVIS still failed to learn an optimum strategy to categorise these stimuli.

Number of stimuli, may not be enough for the implicit system re: W therefore maybe modelling other experimental data, that utilised stimuli with more stimulus dimensions and therefore more possible stimuli may demonstrate the utility of the implicit system of COVIS.

Simulation 3: Medin and Schaffer (1978)

This last simulations will describe COVIS's ability to predict the results from Experiment 2 in Medin and Schaffer (1978). This is another data set that has been widely used to test formal models of categorisation (e.g. Palmeri & Nosofsky, 1995; Smith & Minda, 2000).

The stimuli used by Medin and Schaffer (1978) varied on four binary dimensions (form, size. colour and position) resulting in 16 possible stimuli. In Experiment 2 participants were trained to categorise nine, randomly ordered, stimuli as either Category A or B in a training phase with feedback given according to the category structure in Table 5, until either the participant was able to correctly classify all nine stimuli or each stimulus had been presented 16 times. They were then presented with both the training stimuli and seven novel transfer stimuli to categorise without feedback.

Fitting COVIS to Experiment 2 from Medin and Schaffer (1978)

The version of COVIS needed to model the training period of the experiment was exactly the same as used in the two earlier simulations. However, some modifications were needed to account for the lack of feedback in the transfer task. Although this was simple for the procedural-learning system (merely changing the value of the Obtained Reward to

Table 5. The categorisation structure used in Experiment 2, Medin & Shaffer (1978)

Training Stimuli			Transfer Stimuli	
Category A		Stimulus Number		Stimulus Number
	1 1 1 0	4	1 0 0 1	1
	1 0 1 0	7	1 0 0 0	3
	1 0 1 1	15	1 1 1 1	6
	1 1 0 1	13	0 0 1 0	8
	0 1 1 1	5	0 1 0 1	9
Category B			0 0 1 1	11
	1 1 0 0	12	0 1 0 0	16
	0 1 1 0	2		
	0 0 0 1	14		
	0 0 0 0	10		

0), it was more complicated for the explicit and competition systems as, unlike the procedural-learning system, Ashby et al. (2011) fail to definitively detail how to encompass lack of feedback for these systems.

The lack of feedback affects the explicit system in two ways. Firstly, the perceived reward (Δ_C) and the perceived cost of an error (Δ_E) in Equations 2 and 3 were set to zero during the transfer task. This is because without feedback it is impossible to determine whether the salience of that rule should increase or decrease. Also, without knowing whether the response is correct or not, COVIS is unable to decide whether to use the previous rule or to select a rule on the basis of Equation 7. Therefore, when implementing the model during the transfer task I assumed that it would choose the model based solely on Equation 7.

In the competition system, the problem lies with determining the degree of trust COVIS should place each system as this is determined by the past successes or failures of each system. Therefore, both Equations 16 and 17 were removed and replaced by

$$\theta_E(n+1) = \theta_E(n) \tag{19}$$

Trust in the procedural system, θ_P, remains as given by Equation 18. This means the competition system is able to take advantage of the learning process during training but once the transfer task starts trust in both systems doesn't change.

The model was fitted as in SImulation 2. The parameters determined by the optimisation algorithm are listed in Table 7.

Model Fitting Results

As can be see from Figures 15, COVIS was unable to capture much of the structure of the training stimuli. The transfer task was accounted for better but was still not as good as other single system, exemplar modesl (see Smith & Minda, 2000).

Discussion

COVIS once again failed to predict the pattern of data found in Experiment 2 Medin and Schaffer (1978). However, in this case it seems COVIS was able to learn which

Table 6. COVIS parameter values used in the simulation of Experiment 2, Medin and Schaffer (1978)

System	Parameter	Value
Explicit system	Δ_C	0.0027
	Δ_E	0.0219
	γ	0.9464
	λ	5.3462
	$Z_k(0)$, for $k = 1, \ldots, 4$	0.25
Procedural system	D_{base}	0.20
	α_w	0.6203
	β_w	0.2021
	γ_w	0.0212
	θ_{NMDA}	0.0023
	θ_{AMPA}	0.0101
	σ_{P}	0.0125
Competition	Δ_{OC}	0.01
	Δ_{OE}	0.0439

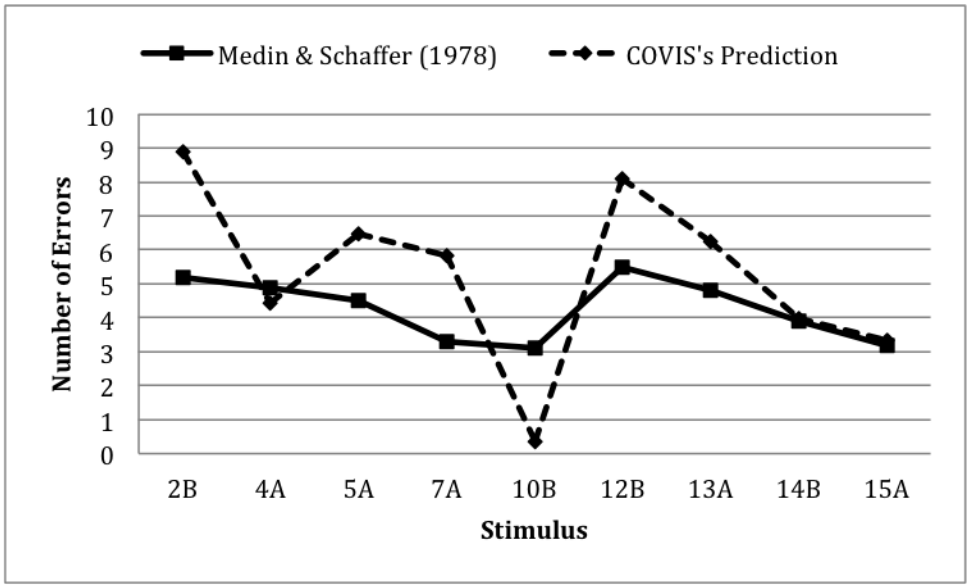

Figure 15. COVIS's prediction vs MS1978 on Training

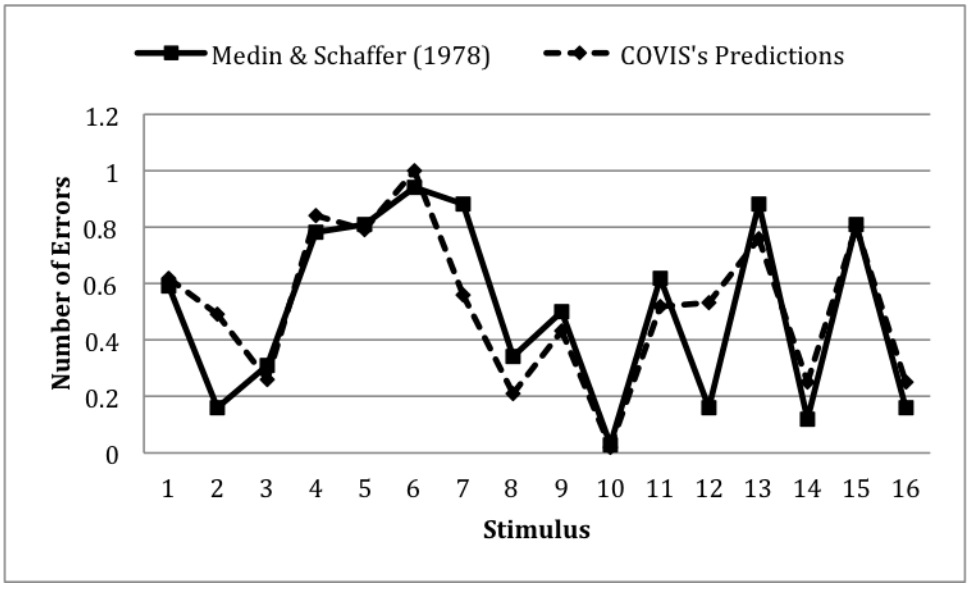

Figure 16. COVIS's prediction vs MS1978 on the Transfer Task

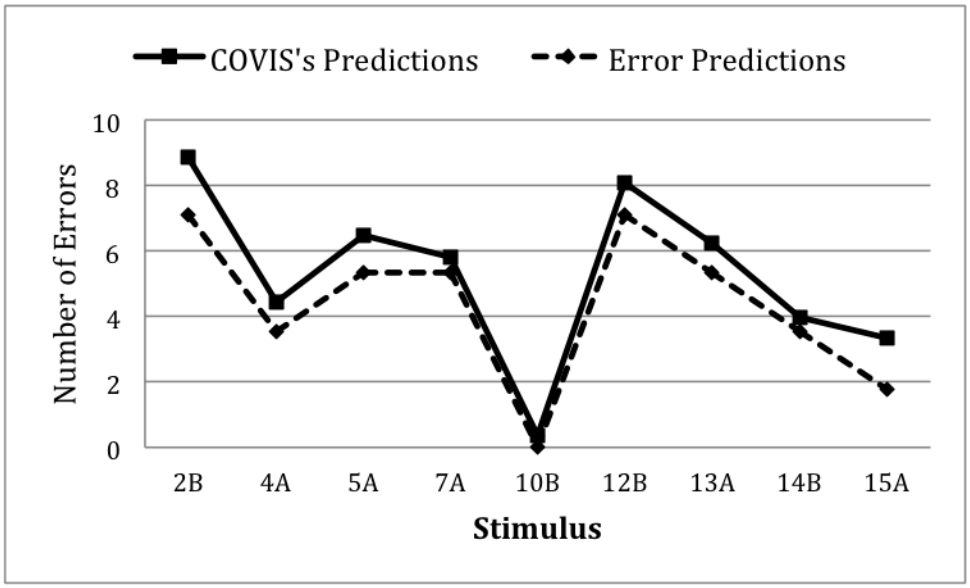

Figure 17. COVIS's predicted error compared with the expected errors based on a one dimensional rule for Medin and Schaffer (1978)

one-dimensional rule was optimum. Looking only at the training stimuli, it is possible to determine the proportion of correct category assigns that would be obtained by using a one-dimensional rule. Rules based on stimulus dimensions 1 or 3 are the most accurate, followed by dimension 4 with the less reliable being rules based on dimension 2.

If the weighted average numbers of errors expected for each category is calculated according to the success that should be expected from each rule, a curve very similar to that predicted by COVIS results, see Figure 17. This indicates that once again, COVIS has only used the verbal system to categorise the stimuli into the appropriate category. Therefore, it seems that once again COVIS assigns category labels solely using the verbal

Table 7. Proportion of correct category assignments expected if a one-dimensional rule was used

Dimension:	1	2	3	4
Category A	4/5	3/5	4/5	3/5
Category B	3/4	2/5	3/4	3/5
Total	7/9	5/9	7/9	6/9

system.

General Discussion

Overall, COVIS appears to be unable to learn to assign stimuli to categories that aren't defined by a simple one-dimensional rule. In Simulation 1, COVIS failed to replicate the ordinal properties of the data in Waldron and Ashby (2001) and both category types were solved using the verbal system, contrary to the assumptions of COVIS. Similarly, in Simulation 2, COVIS failed to capture the properties of the Nosofsky et al. (1994) learning curves in such a way as to suggest it was still only using the explicit system. The failure of COVIS to capture the data from Medin and Schaffer (1978)

1.

2. Future directions: fit simultaneously

References

Ashby, F. G., & Maddox, W. T. (2005, January). Human category learning. *Annual review of psychology*, *56*, 149–78. Available from http://www.ncbi.nlm.nih.gov/pubmed/15709932

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). New York: Cambridge University Press.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*, 349–377.

Helie, S., Paul, E. J., & Ashby, F. G. (2011). Simulating Parkinson's disease patient deficits using a COVIS-based computational model. *Proceedings of International Joint Conference on Neural Networks*, 207–214.

Hélie, S., Paul, E. J., & Ashby, F. G. (2012, August). Simulating the effects of dopamine imbalance on cognition: From positive affect to Parkinson's disease. *Neural networks : the official journal of the International Neural Network Society*, *32*, 74–85. Available from http://www.ncbi.nlm.nih.gov/pubmed/22402326

Kruschke, J. K. (1992, January). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological review*, *99*(1), 22–44. Available from http://www.ncbi.nlm.nih.gov/pubmed/1546117

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004, April). SUSTAIN: a network model of category learning. *Psychological Review*, *111*(2), 309–32.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension

of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*, 352–369.

Nosofsky, R. M., & Palmeri, T. J. (1996, June). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review*, *3*(2), 222–226.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 548–568. Available from `http://psycnet.apa.org/journals/xlm/21/3/548/`

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: current research and theory* (pp. 64–69). New York: Appleton-Century-Crofts.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.

Smith, D. J., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. Available from `http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.26.1.3`

Tharp, I. J., & Pickering, A. D. (2009, June). A note on DeCaro, Thomas, and Beilock (2008): further data demonstrate complexities in the assessment of information-integration category learning. *Cognition*, *111*(3), 411–415.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, *87*(1), 1–8.

Waldron, E. M., & Ashby, F. G. (2001, March). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168–76.

Wills, A. J., & Pothos, E. M. (2012, January). On the adequacy of current empirical

evaluations of formal models of categorization. *Psychological Bulletin*, *138*(1), 102–125.