Running Head: TRAINING TYPE AND CATEGORIZATION

Feedback can be superior to observational training for both rule-based and information-integration category structures.

C. E. R. Edmunds[1], Fraser Milton[2] and Andy J Wills[1]

[1]School of Psychology, University of Plymouth, UK.

[2]Department of Psychology, University of Exeter, UK.

Word Count=8778.

Author Note

C. E. R. Edmunds, School of Psychology, Plymouth University, Drake Circus, Plymouth. PL4 8AA. United Kingdom.

Email: charlotte.edmunds@plymouth.ac.uk

Our thanks go to Angus B. Inkster for his help coding the verbal reports.

Abstract

The effects of two different types of training on rule-based and information-integration category learning were investigated in two experiments. In observational training, a category label is presented, followed by an example of that category and the participant's response. In feedback training, the stimulus is presented, the participant assigns it to a category and then receives feedback about the accuracy of that decision. Ashby, Maddox, and Bohil (2002) reported that feedback training was superior to observational training when learning information-integration category structures, but that training type had little effect on the acquisition of rule-based category structures. These results were argued to support the COVIS dual-process account of category learning. However, a number of non-essential differences between their rule-based and information-integration conditions complicate interpretation of these findings. Experiment 1 controlled, between category structures, for participant error rates, category separation, and the number of stimulus dimensions relevant to the categorization. Under these more controlled conditions, rule-based and information-integration category structures both benefitted from feedback training to a similar degree. Experiment 2 maintained this difference in training type when learning a rule-based category that had otherwise been matched, in terms of category overlap and overall performance, with the rule-based categories used in Ashby et al. These results indicate that differences in dimensionality between the category structures in Ashby et al. is a more likely explanation for the interaction between training type and category structure than the dual-system explanation they offered.

KEYWORDS: COVIS, categorization, implicit, explicit, feedback.

Ashby and Maddox (2011) stated that many researchers now assume multiple systems are involved in category learning. To the extent that this claim is accurate, it is down in no small part to the behavioral dissociations reported by Ashby, Maddox and colleagues. These studies tend to find a differential effect of a manipulation on the learning of two types of category structure: rule-based and information-integration. Ashby and Maddox (2011) argue that these dissociations are predicted by one particular dual-system model of category learning, COVIS (COmpetition between Verbal and Implicit Systems; Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby, Paul, & Maddox, 2011), which assumes the existence of two competing systems of category learning. The strength of the case for COVIS is, of course, not a function of the number of dissociations that have been reported, but rather of the number that prove to be reliable and valid. Indeed, there is a growing body of work that casts doubt on the validity or interpretation of a high proportion of these dissociations (e.g. Dunn, Newell, & Kalish, 2012; Newell, Dunn, & Kalish, 2010; Newell, Moore, Wills, & Milton, 2013; Stanton & Nosofsky, 2007, 2013). In light of this accumulation of critiques, it becomes particularly important to assess the remaining dissociations. In the current article, we report a re-examination of an influential dissociation reported by Ashby, Maddox, and Bohil (2002), which has not been previously re-examined.

Ashby et al. (2002) compared the effect of observational and feedback training on categorization performance. On each trial in observational training, participants were shown the correct category label, followed by the stimulus, and then made a classification response. In feedback training, participants were shown the stimulus, made a classification response and then received feedback on the accuracy of that response. The stimuli were lines that varied in length and orientation. Two different category structures were considered: a unidimensional rule-based structure, such as Figure 1(a), and a two-dimensional diagonal information-integration structure, such as Figure 1(b). Ashby et al. found that participants' performance in the unidimensional rule conditions were similar regardless of training type, whereas participants in the information-integration conditions were less accurate with observational training than those with feedback training. They argued that these findings support the COVIS model of category learning (Ashby et al., 1998, 2011).

Figure 1 about here

Ashby et al.'s (2002) dissociation is predicted by COVIS because the model assumes that rule-based and information-integration category structures are most effectively learned via dissociable neural systems that utilize feedback differently (Ashby et al., 1998). The Verbal System relies on explicit, logical reasoning and excels at learning rule-based categories by testing simple verbal rules such as "short lines belong to Category A and long lines belong to Category B", such as Figure 1(a), or conjunctive rules such as "large, horizontal lines belong to Category A, otherwise they belong to Category B", illustrated in Figure 1(c). The Verbal System operates via a process of hypothesis generation and testing that utilises working memory to maintain representations of the stimulus and the current rule long enough to learn regardless of the order in which the information is presented (Ashby et al., 2002). Consequently, as found by Ashby et al., COVIS predicts that training type should have little effect on the learning of rule-based categories. In contrast, the Implicit System integrates information from the multiple stimulus dimensions pre-decisionally and associates this representation with a particular motor response. The Implicit System is proposed to be responsible for learning "information-integration" category structures, illustrated in Figure 1(b), where the perceptual boundary between the categories is difficult or impossible to describe verbally and therefore cannot be optimally learned by the verbal system (Ashby et al., 1998). The Implicit System is hypothesised to be sensitive to how feedback is presented. It relies on unexpected reward to learn, so should learn more effectively when feedback follows a response than when the category label precedes the response (Ashby & Maddox, 2003). This means that COVIS predicts, as found by Ashby et al. (2002), that learning of information-integration categories should be impaired with observational training relative to feedback training.

In terms of COVIS, the critical difference between the rule-based and information-integration categories is that the former structure is readily verbalizable whereas the latter is not (Ashby et al., 1998). This is because verbalizability determines which system is responsible for optimum responding. Therefore, an ideal test of COVIS's predictions about the effect of training type on category learning should vary verbalizability while holding other potential confounds constant. However, Ashby et al.'s (2002) study contained three superfluous factors that varied between the rule-based and information-integration categories. First, the number of dimensions required to accurately learn each category varied: the information-integration structure required participants to utilise both stimulus dimensions, whereas the rule-based category structures only required one. Single-dimension classification has been shown to sometimes require less cognitive resources (as indexed by the effects of concurrent load and time pressure) than multi-dimension classification (Milton, Longmore, & Wills, 2008; Wills, Milton, Longmore, Hester, & Robinson, 2013). Therefore, training type may be less critical in the rule-based conditions than the information-integration conditions because it is a less demanding category structure.

Second, participants in Ashby et al.'s (2002) first experiment made very few errors in the rule-based conditions, but rather more in the information-integration conditions, raising the possibility that the observed dissociation was the result of a ceiling effect. Ashby et al. partially addressed this possibility by running a second study in which the rule-based structure was made harder to learn by reducing the between-category separation. Although the overall performance of participants decreased, there was still no statistically significant difference between observational and feedback training for rule-based categories under these conditions, supporting Ashby et al.'s interpretation. That being said, performance was marginally better with feedback training compared to observational training. In addition, difficulty was only increased for one of the two combined counterbalance conditions, and there were only five participants per condition. Thus, the lack of a significant difference in Ashby et al.'s Experiment 2 might be attributable to a lack of statistical power.

Third, in both of Ashby et al.'s (2002) experiments, the rule-based structures had lower category separation than the information-integration structures. Category separation is the mean distance between category items as plotted in stimulus space, divided by the within-category variance along the direction of the comparison. Given that differences in category separation were shown by Stanton and Nosofsky (2007) to be responsible for the dissociation in another paper purported to support COVIS (Maddox, Ashby, Ing, & Pickering, 2004), it seems important to control for this factor in future investigations of Ashby et al.'s (2002) dissociation.

Although it is difficult to simultaneously control all three of these factors (number of relevant dimensions, error rates and category separation) while maintaining the essential difference in verbalizability, this goal has been achieved in other COVIS-related studies. Specifically, Filoteo, Lauritzen, and Maddox (2010) in their study of the effects of concurrent load on rule-based and information-integration category learning, employed the category structures illustrated in Figures 1(b) and 1(c). Filoteo et al.'s rule-based structure is a conjunctive rule and so requires participants to be sensitive to both stimulus dimensions. Furthermore, Filoteo et al.'s study establishes empirically that these rule-based and information-integration structures are well matched on participant error rates. They are also closely matched on category separation.

For these reasons, Experiment 1 re-examined the effect of feedback compared to observational training using the category structures utilized by Filoteo et al. (2010). For this experiment, COVIS predicts that feedback training should be superior to observational training for the information-integration structure, but that training type should matter relatively little for the rule-based structure. However, Ashby et al.'s data is also consistent with the hypothesis that feedback is superior to observation for both rule-based and information-integration category structures. This is because the dissociation observed by Ashby et al. may be due to one or more of the superfluous factors for which they did not control (participant errors, category separability, problem dimensionality). Under this latter hypothesis, the current experiment should show a similar feedback advantage for both category structures, because these superfluous factors have been better controlled.

In addition to an examination of response accuracy, we also asked participants, at the end of the experiment, to describe their classification strategies. Not only does previous evidence indicate that reported strategy use can be informative when comparing the effect of feedback and observational training on a probabilistic category learning task (Newell, Lagnado, & Shanks, 2007), but it can also directly assesses whether participants can verbalize the category structure. If the rule-based, conjunction category structure is more verbalizable than the information-integration category structure, then participants should be more successful at describing the underlying structure in the rule-based condition than the information-integration condition. Also, the use of model-based analysis of participants' responses, based around General Recognition Theory (GRT; Ashby & Gott, 1988), is standard practice within experiments inspired by the COVIS model. Although we have some reservations about this procedure, we have presented these analyses to facilitate comparison with other work in this field.

Experiment 1

Method

Participants and apparatus 80 participants (47 female) were recruited from the University of Exeter community and were not rewarded for their participation.

The experiment was run using MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extensions on a MacBookPro with a 15-inch screen.

Design The experiment had a 2 (category structure: rule-based, information-integration) x 2 (training type: observation, feedback) between-subjects, factorial design. 20 participants were randomly assigned to each condition. Category learning was measured by the percentage of correct responses in each block.

Stimuli We used the same stimuli as in the two-dimensional information-integration, Figure 1(b), and rule-based, Figure 1(c), conditions of Filoteo et al. (2010). Each stimulus was a single black line on a white background that varied on two dimensions: line length and orientation. In both conditions, maximum accuracy was 95% as 5% of the stimuli overlapped the optimal category boundary.

Procedure Participants in all conditions were informed that they would be shown a series of lines that varied in length and orientation, that their task was to assign the lines to either Category A or Category B and that approximately half the lines were in each category. They were also told that at the beginning they may have to guess but by the end they should be able to reach high levels of accuracy. They were further informed of the structure of the experiment, the format of the trials, the position of feedback within the trial (which varied between conditions) and the response keys.

The experiment consisted of 10 blocks of 60 trials, with 600 trials in total. Participants assigned stimuli to either Category A (by pressing the 'Z' key) or Category B (by pressing the '/' key). Starting with a training block, the blocks alternated between training and test. This was to provide a measure of performance during learning for both observational and feedback conditions as well as to facilitate comparison with Ashby et al. (2002). The training trials of the feedback learning conditions consisted of displaying the stimulus for 500ms, followed by a blank screen for 500ms, followed by a self-paced classification response. Finally the correct category label was displayed for 500ms. In the observational learning condition training trials consisted of first displaying the correct category label for 500ms, followed by a blank screen for 500ms, followed by the stimulus for

500ms to which the participant made a self-paced response. The test trials in both feedback and observational training conditions included no information about the correct category assignment and consisted of a stimulus displayed for 500ms followed by a self-paced response. The inter-trial interval in all conditions was 500ms.

At the end of the experiment, participants were presented with a questionnaire that asked them to describe whether they had a specific strategy when classifying the items and, if so, to describe it, using either words or pictures.

Data archiving The trial-level raw data are archived at `www.willslab.co.uk/exe201201/` with md5 checksum `3a294887e14de59dc09bab76c27a9162`[1].

Results

Following Ashby et al. (2002), analyses were conducted on the final test block of the data from all participants. Conducting the analyses across all test blocks led to the same conclusions, as did excluding participants failing to reach 50% on the final block (the analysis method and exclusion criterion applied by Filoteo et al., 2010). Figure 2 shows mean accuracy for each condition in just the last test block, Figure 2(a), and across all test blocks, Figure 2(b).

Figure 2a) and 2b) about here

An ANOVA revealed a significant main effect of training type, $F(1, 76) = 7.68$, $\eta^2 = 0.09$, $p = .007$, but not of category structure, $F(1, 76) = 1.89$, $\eta^2 = 0.02$, $p = .175$. Hence, participants learned more in the feedback training condition than in the observational training condition when learning both rule-based and information-integration categories. The interaction between training type and category structure was also

[1]Publication of an MD5 checksum allows the reader to independently confirm that the raw data in the archive is unchanged.

non-significant, $F(1, 76) = 0.058$, $\eta^2 = 0.00$, $p = .811$.

Bayesian analysis The standard statistical analyses above indicate that, unlike in Ashby et al. (2002), there appears to be no difference between rule-based and information-integration categories in the effect of training type on learning. However, in null-hypothesis significance testing, non-significant results are ambiguous; they could either be due to insufficient statistical power or due to the null hypothesis being correct (Dienes, 2011). As the interaction between training type and category structure formed the basis of the conclusions drawn by Ashby et al. (2002), it is important to determine whether the reason the current study failed to find an effect was because it lacked power. One way of determining this is to calculate Bayes Factors for the relevant comparisons (Dienes, 2011). Briefly, if the Bayes Factor is over three then the experiment has found evidence for the experimental hypothesis whereas if the Bayes Factor is less than a third, the experiment finds evidence for the null hypothesis (Jeffreys, 1961). A Bayes Factor of one indicates that the evidence is exactly neutral with respect to the experimental and null hypotheses (Dienes, 2011). Values between a third and three are typically interpreted as indicating that the experiment was not sensitive enough and no conclusions can be drawn.

To calculate the Bayes Factor for the interaction between category structure and learning type we followed the procedure recommended by Dienes (2011). This requires the expected average difference between the two differences to be specified. In Ashby et al. (2002), the observed mean difference of differences between the information-integration conditions in Experiment 1 and the rule-based conditions in Experiment 2 were approximately 15%, and we used this figure in our analyses. This cross-experimental difference was used as the rule-based structure in Experiment 2 was better controlled for differences in overall error rates. Following the recommendations of Dienes (2011), we assumed a normal distribution around this mean with standard deviation of half the mean (i.e. 7.5, representing the experimental hypothesis that differences as small as zero are unlikely). These calculations result in a Bayes Factor of 0.18. As the Bayes Factor is less than a third, it indicates that the data provides support for the null hypothesis, i.e. that there is no difference between rule-based and information-integration category learning in the effect of varying training type. These conclusions held even if the expected average difference between the rule-based and information-integration conditions was underestimated by up to a third of that reported by Ashby et al.

State-trace plot The analyses above indicate that there are no differences between the acquisition of rule-based and information-integration categories. However, these analyses do not consider the qualitative pattern of learning throughout the experiment. As the key conceptual claim of COVIS is that there are two mechanisms of learning, it could be argued that these analyses have failed to identify multiple systems only because the difference in learning between training types just happened to be the same for rule-based and information-integration learning by the end of training. To examine the validity of this claim, we used state-trace analysis (Bamber, 1979; Loftus, Oberg, & Dillon, 2004), which has previously been used with great success on this type of category learning data (Newell et al., 2010; Dunn et al., 2012).

State-trace analysis is an alternative to dissociation logic that allows experimenters to determine whether multiple systems are required to explain an experimental result. This is accomplished by drawing a state-trace plot. To do this, two dependent variables, in this case performance on the rule-based and information-integration category structures, are plotted on the x and y axes. Then, a trace is plotted for each training type condition, with each point being the accuracy from each test block. The state-trace plot is then inspected to determine whether the traces are consistent with a single- or multiple-system account. If the two traces overlap to form a single monotonic function, then there is an absence of evidence that a multiple-process account is required to explain the observations. If the traces form two monotonic functions, then this is often interpreted as being more supportive of a multi-process account, although the question of what the term "multi-process" means in the context of state-trace analysis has been the topic of recent debate (Dunn, Kalish, & Newell, 2014). In brief, both Yeates and colleagues (Yeates, Wills, Jones, & McLaren, 2012, in press) and Ashby (2014) have identified situations where models typically considered to be single-system accounts can produce two functions on a state-trace plot through variation in a single parameter (specifically, attention weight in the Generalized Context Model, Nosofsky, 1986, and learning rate in the Simple Recurrent Network, Elman, 1990).

Figure 3 about here

From visually inspecting Figure 3, the data from the current experiment forms a single monotonic curve. This suggests an absence of evidence that a multi-process account such as COVIS is required to account for the current results. However, it is worth noting that to conclusively infer this, the plot should be statistically tested for a significant departure from monotonicity.

Model-based analyses The COVIS-based predictions for this data set (see Introduction) are contingent on the assumption that the category type manipulation corresponds to a change in the learning system that controls responding. Practically, this means that there should be more people using the verbal system in the rule-based category conditions than in the information-integration conditions, and vice-versa for the implicit system. Experimental studies within the COVIS framework utilize model-based analysis constructed from GRT (Ashby & Gott, 1988) to examine this assumption. For each participant, this analysis determines the optimum decision boundary in stimulus space that separates the stimuli judged by each participant to be in Category A from those in Category B. Each participant is then assigned a strategy type, such as unidimensional, on the basis of characteristics of their optimum boundary. The assumption that the category type manipulation has resulted in a change of category learning system is argued to be valid if more participants are using the optimum decision bound for the category structure they have been assigned to, such as a diagonal decision boundary in the information-integration conditions, than are using that strategy in the inappropriate category structure, such as a diagonal decision boundary in the rule-based conditions.

The GRT-based analysis determines which of a pre-defined set of decision-boundary models best describes the classification each participant has produced. The set of models considered in this analysis were as follows:

The *unidimensional* models assume that the participant determines a criterion along one of the stimulus dimensions, either orientation or length. They then make a decision about the category membership of each stimulus by comparing the appropriate stimulus attribute with the criterion value. As an example, for length, this corresponds to a rule of the type: 'Assign to Category A if the stimulus is long, or Category B if short'. The unidimensional models have two parameters: the value of the criterion and the variance of internal (criterial and perceptual) noise.

The *conjunction* model assumes that the participants make two judgements, one for each stimulus dimension, and then combine these to make a judgement about category membership. The conjunction rule in the current analysis was of the type: 'Assign to Category A if the stimulus is short and upright, otherwise assign to Category B'. The conjunction model had three parameters: the two criterion values and internal noise.

The *General Linear Classifier* (GLC) model assumes that the decision boundary between the categories can be described by a straight line that can vary in gradient and intercept. The unidimensional models are therefore special cases of the GLC model. The GLC model has three parameters: the intercept and slope of the decision bound, plus noise.

The *random* model assumes that participants are responding randomly; it has no parameters.

For each participant, the best fit of each of these models was calculated, and the best-fitting model selected using Akaike's information criterion (Akaike, 1974). The results from this analysis, which was performed using the `grt` package in the R environment (Matsuki, 2014), are reported in Table 1. Within the COVIS framework, the unidimensional and conjunction models are considered to represent explicit, rule-based strategies, while the GLC is considered to represent an implicit, information-integration strategy.

Table 1 about here.

In ordinal terms, the results of this analysis are consistent with the intended effects of the experimental manipulation, as seen through the lens of the COVIS model and GRT-based model analysis. Specifically, the proportion of participants best fit by a conjunction model is higher in the rule-based condition than the information-integration condition, and the proportion of participants best fit by the GLC model is higher in the information-integration condition than in the rule-based condition.

It is perhaps not particularly surprising that some participants are best fit by a unidimensional model, as a single-dimension strategy can optimally achieve approximately 75% accuracy in both the rule-based and the information-integration conditions. From a COVIS perspective, it is not particularly problematic if some participants in the rule-based condition are in fact employing a unidimensional rule, as this is still a rule-based strategy and readily verbalizable. It is potentially more problematic from a COVIS perspective that there are a reasonable proportion of participants best fit by unidimensional models in the information-integration condition, potentially implying the presence of significant rule-based responding in these conditions. A similar result was observed in Ashby et al. (2002), although the proportion is higher in the current study. The presence of unidimensional responders in an information-integration condition is typically accommodated within COVIS by assuming that some participants have not yet transitioned from the explicit system to the implicit system. The lower proportion of participants best fit by unidimensional models in Ashby et al. (2002) may be due to the fact that Ashby et al., in their modelling of their information-integration condition, constrained the GLC model to have the gradient and intercept defined by the category structure. This constrained version of the model has just one parameter, while the unconstrained version we employed has three parameters. In an AIC model-selection procedure, reducing the number of free parameters of a model will, other things being equal, increase the proportion of participants best fit by that model. Somewhat surprisingly, Ashby et al. state that they employed the unconstrained version of the GLC in their fits of their rule-based condition. This difference in fitting procedure between experimental conditions seems odd, and may have contributed to the higher proportion of unidimensional classifiers in their rule-based conditions compared to their information-integration conditions.

In summary, the model-based procedures that are standard in this field broadly support the supposition that participants in the rule-based conditions classify the stimuli differently to participants in the information-integration conditions. The fact that a conjunction model best fits more participants in the rule-based condition than the information-integration condition, and a GLC model best fits more participants in the information-integration condition than the rule-based condition, is broadly consistent with the predictions of the COVIS model. Of course, what is not consistent with the COVIS model is that, despite these differences, there is no difference in the size of the feedback advantage in the rule-based and information-integration conditions.

Although seldom reported within the COVIS literature, it is also informative to look at the performance of the best-fitting model relative to the performance of the competing models. If the winning model performs much better than its competitors, we can be fairly confident that this model provides the best description of the participant's behavior, from among the pre-specified alternatives. On the other hand, if the competing models perform almost as well as the winning model, our confidence that the winning model provides the best description should probably be lower.

One principled way of evaluating the model-based analysis is by calculating the normalized probability that a conjunction model is preferred to the GLC for each participant (or vice versa). This is done by calculating the Akaike weight, $w_i(AIC)$, for each model for each participant (Wagenmakers & Farrell, 2004). This is defined as the probability that model i is the best, in terms of minimising the Akaike information criterion, given the data and the set of competing models. From the Akaike weights, the normalized probability that Model i is to be preferred over Model j is calculated using

$$\frac{w_i(AIC)}{w_i(AIC) + w_j(AIC)} \tag{1}$$

where $w_i(AIC)$ and $w_j(AIC)$ are the Akaike weights for models i and j respectively.

For the rule-based category structure conditions the probability of the 'best' model being a conjunction, rather than the GLC, is 0.635 in the feedback training condition and 0.668 in the observational training condition. This provides additional support that participants are genuinely using orthogonal decision boundaries to make decisions. However, for the information-integration category structures the probability of the best model being the GLC, rather than a conjunction, is much lower: 0.297 for the feedback training condition and 0.382 for the observational training condition. Clearly, confidence in the results of GRT-based model fitting in the information-integration conditions should be low. We would be interested to see comparable information for Ashby et al. (2002), or any other COVIS-relevant study, and suggest this or a similar measure be included in future research.

Verbal report analysis An alternative explanation of these findings from within the COVIS framework might be that the majority of participants in both the rule-based and information-integration category structure conditions were using the implicit system. It is possible that by increasing the number of relevant dimensions in the rule-based structure, participants found this too difficult and so resorted to using the implicit system. To investigate this possibility we examined the strategies reported by participants as summarized in Table 2.

The verbal reports were independently coded by two of the authors (CERE and

AJW) and any discrepancies that were not due to human error were easily resolved through discussion. First, each verbal report was examined to determine whether the participant had reported an explicit categorisation strategy or not. The inter-rater reliability for this was perfect, $\kappa = 1$, $p < .001$. Second, the available strategy descriptions were sorted into groups of three main kinds: unidimensional, two-dimensional and miscellaneous.

Participants were placed in the *unidimensional length* or *unidimensional orientation* groups if they described categorizing stimuli based solely on line length or line orientation respectively.

Participants were placed in the *conjunction* group if they used both stimulus dimensions and described categorizing stimuli using a logical conjunction rule such as 'short, upright lines were in Category A, otherwise they were in Category B.'

Participants were placed in the *information-integration* group if they described attempting to make the stimulus dimensions commensurable, such as 'Stimuli for which the line was longer than it was upright should be assigned to category A' or if they said anything that could be reasonably interpreted as a statement that they based their classification on overall similarity. Note that overall similarity descriptions are commonly found in other studies, not within the COVIS-framework, in which we have elicited verbal reports (e.g., Wills et al., 2013).

Participants were placed in the *two-dimensional* group if they described using both stimulus dimensions but with descriptions that were too unclear to be assigned to more specific categories.

All remaining participants were assigned to the *other* group, which included participants whose descriptions were too vague to be assigned to another group.

Inter-rater reliability for strategy assignment was high, $\kappa = .813$, $p < .001$, with the majority of discrepancies appearing to be due to human error in applying the strategy definitions, rather than any inherent ambiguity in the definitions themselves (as all discrepancies were rapidly resolved by reference to the strategy descriptions). There were no significant differences between all conditions in the number of participants who did not report a strategy, $\chi^2(1) = 0.12$, $p = .730$. With respect to the types of strategy reported, there are very different patterns of responding between the rule-based and information-integration category structure conditions. For the rule-based conditions, although there is clearly some variability, the modal strategy correctly described the conjunction structure. In addition, none of the participants in these conditions reported using an overall similarity or information-integration strategy, and only 20.1% reported using unidimensional strategies.

In contrast, no participant in the information-integration category structure conditions reported any strategy that could be interpreted as describing the structure of the information-integration category they had been presented. In these conditions, participants were equally likely to report a unidimensional strategy as they were to report a conjunction rule, although strategies employing both dimensions were the majority indicating a sensitivity to the fact that both dimensions were relevant. This summary is supported by the fact that the number of participants in the rule-based category conditions who reported the optimal strategy for the categorization problem they had been presented (44.8% of the people who reported strategies) was significantly different from those in the information-integration conditions who identified the correct strategy (0% of the people who reported strategies), $\chi^2(1) = 15.20$, $p < .001$.

In sum, although participants found neither category structure trivial to verbalize, participants in the rule-based category structure conditions were more able to verbalize the underlying category structure than those in the information-integration conditions. Thus, these analyses largely support the assertion that the rule-based category structure is more readily verbalizable than the information-integration category structure.

Table 2 about here.

Comparing model-based analyses with verbal reports The model-based analyses and verbal reports used here are complementary approaches that both aim to determine how participants are completing the task. However, from the summaries of these analyses above, it appears that they are partially inconsistent with each other. To examine the degree of correspondence between these approaches, we compared the strategy each participant was assigned using the model-based analysis with the one they reported using after the experiment (Table 3).

As can be seen, for the rule-based strategies (unidimensional, two-dimensional and conjunction) the model-based analyses and verbal reports match reasonably well. This is not the case for the GLC and reports of implicit or overall similarity responding; all participants that were assigned to the GLC strategy in the model-based analysis reported using an explicit rule-based strategy. One possible explanation for this disparity is that participants were using an implicit, GLC based, strategy but were unable to describe it correctly. Although, this may be unsurprising given that it is implicit, it seems unlikely given that in previous, different but related, work participants were able to report this type of strategy (Wills et al., 2013). Alternatively, it may be that the GLC is more inclusive than the other models, and so results in participants that are using a rule-based strategy being assigned to the GLC merely because they could not be assigned to another type of strategy. This later hypothesis is supported by the Akaike weight for the GLC; this model wins by a much lower margin than the others (see model-based analysis section).

Table 3 about here.

Discussion

Ashby et al. (2002) reported, as predicted by COVIS, that performance with feedback training was superior to observational training when learning an information-integration category structure, whereas for a unidimensional rule-based category they found that these training types resulted in comparable performance. In contrast, we found that learning performance in Experiment 1 was better with feedback training than observational training to a similar degree for both category structures. The Bayesian Analysis verifies that there is truly no difference in learning performance between the two category structures. This pattern of performance is not consistent with the claim that there are two systems of category learning that are differentially affected by training type. The state-trace analysis shown in Figure 3 also does not provide any evidence for a dual-system approach. It consists of a single, monotonically increasing curve, which is interpreted as evidence that performance in this experiment can be described by a single system of category learning.

COVIS could encompass the pattern of performance found in Experiment 1, if participants resorted to using the implicit system for both category structures. However, this hypothesis is not supported by the verbal report analysis that found that participants were equally likely to be able to report a strategy in all conditions, but that fewer participants were able to describe the optimal strategy in the information-integration conditions than in the rule-based conditions. Similarly, the model-based analysis indicates that the conjunction model best fits more participants in the rule-based condition than the information-integration condition, and a GLC model best fits more participants in the information-integration condition than the rule-based condition. Therefore, the results of Experiment 1 appear inconsistent with COVIS.

Experiment 2

Ashby et al. (2002) found an interaction between training type and category structure. They argued that this pattern of results supported COVIS. However, Ashby et al. included several confounds in their design that complicates interpretation of their results: the number of stimulus dimensions relevant to categorization, category separation and error rates. When these were controlled for in our Experiment 1, feedback training was superior to observational training when learning both rule-based and information-integration categories—a pattern of results not consistent with COVIS. The key difference between Experiment 1 and Ashby et al.'s findings is the appearance of a feedback training advantage for rule-based categories. Experiment 2 of the present paper aimed to determine which of the controlled for confounds might have resulted in this difference in the effects of training type.

The number of dimensions relevant to classification seemed to be the most likely cause of the difference between our Experiment 1 and Ashby et al. (2002). This is because Ashby et al. (2002) manipulated category separation and error rates in a second experiment and still did not find a statistically significant difference in performance due to training type. Therefore, in Experiment 2 of the current paper, to discriminate dimensionality from the other factors, the number of relevant dimensions in the category structure were maintained whilst category separation and error rates were varied. Category separation was increased. Error rates were reduced by scaling the length dimension to increase perceptual discriminability along that dimension and on each trial the stimulus, category label and inter-trial interval were increased to 1000ms.

If increased error rates or reduced category separation are the cause of the difference in learning rule-based categories between our first experiment and Ashby et al. (2002) then the difference between training type should disappear in this experiment. However, if the locus of the difference is the number of relevant dimensions for the rule then the advantage for feedback training over observational training should remain.

Method

Participants and apparatus 40 participants (10 male) were recruited from the Plymouth University paid pool and were paid £8 for their participation.

The experiment was run on a desktop computer on a 21.5-inch screen using MATLAB 2012b with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Design The experiment had 2 between-subjects conditions (training type: observation, feedback). 20 participants were randomly assigned to each condition. Category learning was measured by the percentage of correct responses in each test block.

Stimuli This version of the experiment still utilised a conjunction category structure. However, the category structure was altered to make learning easier (Figure 4). To generate the category structure, four sets of points, 300 from the Category A distribution and 100 each from the other three, were randomly selected from bivariate normal distributions defined using the parameters listed in Table 4. Any points that were over 2.25 standard deviations away from the mean of the distribution in the direction of the category boundary were resampled. Then, as Experiment 1 indicated that the orientation of the line stimuli appeared more salient than line length to participants, the distribution was scaled so that the lines varied between 20 and 350 points in arbitrary units.

Table 4 and Figure 4 about here.

Procedure The only change to the procedure of Experiment 1 was that inter-trial interval, as well as the duration of the stimulus and category label presentation were increased from 500ms to 1000ms.

After the experiment, participants again completed a questionnaire to identify the strategy they used. The format of this varied slightly from the one used in Experiment 1, based on our experience of coding the Experiment 1 responses, and in an attempt to elicit clearer descriptions. They were asked to “Imagine that another participant was asked to complete the experiment exactly as you did. What instructions would you give them so that they could exactly copy your pattern of responding? Please try to be as precise as possible.”

Data archiving The trial-level raw data are archived at `www.willslab.co.uk/ply27/` with md5 checksum `7beb8e5354548f9852a5412b8ac0dbd9`.

Results

Following Ashby et al. (2002) and Experiment 1, analyses were conducted on the final test block of the data from all participants. Conducting the analyses across all test blocks led to the same conclusions. No participant failed to reach 50% accuracy by the final test block. Figure 5 shows mean percentage accuracy in each condition for all test blocks.

Overall performance, as expected, was higher than for the participants in the rule-based condition in Experiment 1. An ANOVA revealed a statistically significant effect of training type, $F(1,38) = 4.61$, $\eta^2 = .108$, $p = .038$. Hence, as in Experiment 1, participants learned consistently more in the feedback training condition than in the observational training condition (see Figure 5).

Figure 5 about here.

Model-based analysis The proportions of participants using each model in each condition are in Table 5. From this we can see that the majority of participants in both conditions have been identified by the analysis as using either the correct conjunction strategy or another rule-based one. This supports the hypothesis that the participants are using an explicit, rule-based strategy. However, the proportions of participants in each condition that were assigned to the correct conjunction strategy are statistically different, $\chi^2(1) = 5.63$, $p = .018$. This indicates that participants were more successful at determining the underlying category structure in the feedback training condition than in the observational training condition.

Table 5 about here.

We also looked at the performance of the best-fitting model relative to the performance of the competing models in terms of the Akaike weights. For the participants in the feedback training condition the mean normalized probability of using a rule-based strategy compared to information-integration or random strategies is 0.931, whereas for the observational training conditions the normalized probability is 0.770. This demonstrates that, as expected, participants are most likely to use rule-based strategies in these rule-based conditions and that these strategies were clear winners.

Verbal reports The verbal reports were independently coded by one of the authors (CERE) and an independent rater (ABI). Any discrepancies that were not due to human error were easily resolved through discussion.

Inter-rater reliability for judging whether or not each participant reported a strategy was high, $\kappa = .844$, $p < .001$, whereas judgements as to which strategy they were reporting were reasonable, $\kappa = .595$, $p < .001$. The majority of discrepancies were due to different interpretations of how participants might be expected to describe a conjunction category structure. The coded strategies described by participants are shown in Table 6.

There were no significant differences between conditions in the number of participants who did not report a strategy, $\chi^2(1) = 0.36$, $p = .548$. There was also no significant difference between conditions in those who reported the correct conjunction category, $\chi^2(1) = 0.96$, $p = .327$. Therefore, participants in both conditions were capable of not only coming up with a strategy, but the majority were also able to correctly describe the category structure.

Table 6 about here.

Comparing model-based analyses with verbal reports As before, we also looked at the degree of correspondence between the verbal reports given by participants and the model that best fit their responses as determined by the model-based analysis (Table 7).

In this experiment, the verbal reports matched the model-based analysis reasonably well; the majority of participants that reported using a conjunction strategy were also assigned to this in the model-based analysis. Furthermore, as might be expected in learning a rule-based category structure, no participants reported using implicit or overall similarity responding or were best described, in the model-based analysis, by the GLC model.

Table 7 about here.

Discussion

The key difference between Experiment 1 and Ashby et al. (2002) was the appearance of an advantage for feedback training over observational training when learning a rule-based category structure. Experiment 2 aimed to determine which of the factors that varied between these two experiments was responsible for this difference. To do this, Experiment 2 compared performance with feedback and observational training when learning a two-dimensional category, with reduced error rates and increased category separation compared with the category structure used in Experiment 1. Under these conditions, the advantage of feedback training over observational training remained. In addition, the model-based and verbal reports indicate that the majority of participants in both conditions were able to use and verbally describe a conjunction strategy. This indicates that the interaction between training type and category structure in Ashby et al. (2002) appears to be due to differences in dimensionality between the category structures.

General Discussion

Ashby et al. (2002) reported that feedback training was superior to observational training for an information-integration category structure, but that the two training types were comparable for a rule-based category structure. This dissociation has widely been taken as support for the COVIS dual-process theory of category learning (Ashby et al., 1998, 2011) and is the most cited, un-critiqued behavioral support for this model. According to the COVIS framework, the critical manipulation in Ashby et al. (2002) is that rule-based category structures are easily verbalizable, while information-integration categories are not and that this results in participants learning these two types of category structures using different category learning systems. These two systems incorporate feedback differently, therefore accounting for the Ashby et al. findings. However, there were several non-essential differences between the category structures used by Ashby et al., which casts doubt on whether verbalizability is the key factor in eliciting a differential effect of training type on learning performance.

In Experiment 1, we successfully maintained the between category structure difference in verbalizability while matching them for (a) the number of relevant stimulus dimensions, (b) category separation, and (c) overall performance. We did this by combining the procedures of Ashby et al. with two-dimensional category structures adopted from more recent work in the COVIS framework (specifically Filoteo et al., 2010). Once these extraneous factors were controlled for, the category structure by training type interaction found by Ashby et al. did not appear; learning of both category structures was better with feedback training than observational training. Experiment 2 also found a training type difference in learning the two-dimensional rule-based structure when this structure was broadly matched, in terms of category overlap and overall performance, with the rule-based structures used by Ashby et al. This indicates that the appearance of a differential effect of training type on rule-based learning in these experiments appears to be due to the two-dimensional nature of the conjunction structure; these experiments demonstrated an advantage for feedback training over observational training for not only information-integration categories, but also for two-dimensional rule-based categories.

Alternative explanations

Our findings have implications for the COVIS theory of category learning because they are not predicted by COVIS. In Experiment 1, COVIS predicts a greater feedback advantage for the information-integration structure than the rule-based structure, but both conditions benefit from feedback training to a similar degree. In Experiment 2, COVIS does not predict a feedback advantage, yet one is observed. How, then, might the results of both Ashby et al. (2002) and the current paper be explained?

First, we need to explain why feedback training is superior to observational training. Any theory that presumes learning is driven by prediction error (see e.g. Wills et al., 2009, for a review) should be able to accommodate this result because, in observational training, there is nothing to predict. The ALCOVE model (Kruschke, 1992) is one of several possible category learning models in which learning is driven by prediction error, as is the striatal pattern classifier (Ashby & Waldron, 1999) that forms the basis of Ashby's explanation of why a feedback advantage is sometimes observed.

Second, we need to explain why a benefit of feedback training is sometimes not observed. One possibility is that such findings represent absence of evidence rather than evidence of absence. In Ashby et al.'s first experiment, performance on the harder, observational, training condition is close to ceiling, potentially obscuring the effect. In addition, Ashby et al. report a significant feedback advantage for the unidimensional category structure in the first test block (Ashby et al., 2002, p. 673), which smoothly reduces throughout training until it disappears in the final block (Ashby et al., 2002, Figure 3). Ashby et al.'s conclusions are based on the final block. In Ashby's second experiment, there is a numerical trend in the direction we predict, sample sizes are small, and only one of the two counterbalance conditions were below ceiling. Thus, one possibility is that feedback is always advantageous in rule-based category learning, but that some experiments fail to reveal this due to methodological issues.

Another possibility is that the feedback advantage is genuinely absent for single-dimension rule-based category structures, or at least much smaller than it is for multi-dimensional category structures (rule-based or otherwise). Although further research would be required to make this claim securely, it is interesting to speculate how such an effect might be explained if it were to be confirmed. One possibility is that the size of the feedback advantage is related to how effortful the classification is. Dimensional Summation theory (Milton & Wills, 2004) predicts that single-dimension classification is less effortful than multi-dimensional classification, and this prediction has been supported in multiple studies (e.g. Milton et al., 2008; Wills et al., 2013).

In summary, COVIS predicts that there should be an interaction between training type and category structure, with a smaller difference between training types when learning a readily verbalizable category structure compared to a hard to verbalize one. However, the available evidence (from both Ashby et al. and the current studies) indicates that the pattern of performance on these tasks might be better explained by an interaction of training type and the number of dimensions relevant to classification. Of course, these experiments have not completely disentangled verbalizability from dimensionality. In order to do this, one would have to examine the effect of training type on a unidimensional, difficult to verbalize category. This would be difficult as it is hard to conceive of a unidimensional category structure that would be hard to verbalize without redefining what is meant by a stimulus dimension.

More generally, although there is reasonable support for the idea that providing an opportunity for error improves learning (Grimaldi & Karpicke, 2012; Kornell, Hays, & Bjork, 2009; Potts & Shanks, 2013), such an effect is not always seen even in multidimensional category structures (Newell et al., 2007) and, in some memory tasks, the effect is even reversed (Haslam, Hodder, & Yates, 2011). Neither COVIS, nor our alternative explanation, fully captures these results. Further empirical work is required to clearly identify the conditions under which feedback training is superior to observational training.

Dimensionality

As discussed above, it seems likely that it is the problem dimensionality, rather than the problem verbalizability, that drives the results of Ashby et al. (2002) and the current paper. The comparison of a unidimensional rule-based category structure with a 45-degree rotation of that structure in stimulus space has formed the basis of a large number of experiments by Ashby and colleagues. The comparison is initially appealing, because the two structures are in various formal senses identical (e.g. an optimal classifier performs equally well on both structures), yet one is easy to verbalize while the other is hard to verbalize. However, the two structures are not matched on the number of psychological stimulus dimensions relevant to the classification. This raises the broader question of whether a failure to control problem dimensionality underlies other apparently COVIS-supporting dissociations.

A reviewer suggested that dimensionality is unlikely to be driving the difference of our results and those of Ashby et al. (2002) on the basis that pigeons find the two problems equally difficult (Smith et al., 2011), the implication being that if a nonverbal species finds these two problems equally difficult then it must be the verbalizability of the problems rather than their dimensionality that is important. However, even in nonverbal species, a necessary condition of a unidimensional problem being easier than a two-dimensional problem is that the stimulus dimensions are psychologically separable. Without separability, there is no meaningful psychological sense in which the two problems differ in dimensionality. Smith et al. provide no compelling evidence that their stimuli are separable for pigeons.

Another possible response to our claim that dimensionality is the critical factor is to point out that many of the more recent COVIS-supporting dissociations make use of a two-dimensional rule-based structure, thus equating problem dimensionality between rule-based and information-integration problems (e.g. Maddox, Bohil, & Ing, 2004; Maddox, Filoteo, Hejl, & Ing, 2004; Maddox, Filoteo, & Lauritzen, 2007; Maddox & Ing, 2005; Maddox, Love, Glass, & Filoteo, 2008; Zeithamova & Maddox, 2006) and dissociations, predicted by COVIS, still emerge. However, this evidence is not, perhaps, as compelling as it first appears and in recent years it has attracted substantive critiques on a variety of bases from separate labs (e.g Dunn et al., 2012; Newell et al., 2010, 2013; Stanton & Nosofsky, 2013; Zaki & Kleinschmidt, 2013). Our explanation is, therefore, entirely compatible with the existing evidence.

Model-based analysis

Another interesting question raised by this research pertains to the limitations of the GRT informed model-based analysis which is ubiquitously used in analysing experiments within the COVIS framework. This model-based analysis aims to determine how participants are approaching the categorization task, and from this make inferences as to which system is guiding responding. This model-based analysis is commonly interpreted by Ashby, Maddox and colleagues to demonstrate that the category structure factor has successfully manipulated the learning system in control of responding if, for each category structure condition, more participants are assigned the correct strategy than the one appropriate for the other condition. The current work found this between-condition shift in strategies. However, the current work also utilized verbal reports and a state-trace analysis, which, although consistent with each other, are not consistent with the model-based analysis or its interpretation as supporting a dual-system approach. Visual inspection of the state-trace plot does not provide any evidence for multiple systems. Similarly, participants in all conditions were equally able to provide verbal reports. In addition, when examining the goodness-of-fit of each type of model in the model-based analysis using Akaike weights, there seems to be a disparity in the confidence the analysis places in the conjunction and GLC models that might also cast doubt on whether an actual switch between systems has taken place. This is obviously not the place for a detailed discussion and investigation of the conditions under which the model-based analysis is useful. However, future work might determine whether this type of model-based analysis is merely ineffective in this study, or whether it is more generally capturing something different than previously thought.

Evidence for COVIS

This paper adds to the growing body of literature that has critiqued the experimental dissociations argued to support COVIS (Newell, Dunn, & Kalish, 2011).

However, it is also important to note that there are a number of other dissociations that provide support for COVIS that have not yet been challenged. For example, switching response location part-way through training has been found to impact learning information-integration categories, while this manipulation does not affect rule-based category learning (e.g. Ashby, Maddox, Glass, O'Brien, & Filoteo, 2010). Deferring feedback has been found to have a similar impact on learning these two types of category structure (Smith et al., 2014). One might also point to imaging studies that show different neural substrates for rule-based and information-integration category learning (Ashby & Maddox, 2011, but see Milton & Pothos, 2011). Clearly, more work is needed to assess the strength of these and other dissociations taken to support COVIS.

Conclusion

In summary, the current paper casts doubt on the interpretation of the dissociation found by Ashby et al. (2002). The current experiments demonstrated an advantage for feedback training over observational training not only for information-integration categories, but also for two-dimensional rule-based categories. Therefore, category structure dimensionality, rather than verbalizability, appears to be the key factor driving the appearance of an interaction between category structure and training type in the original study. This paper, therefore, adds to the growing literature (e.g. Dunn et al., 2012; Newell et al., 2010, 2013; Stanton & Nosofsky, 2007, 2013) that casts doubt on the validity or interpretation of the experimental evidence for the COVIS model of category learning.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723.

Ashby, F. G. (2014). Is state-trace analysis an appropriate tool for assessing the number of cognitive systems? *Psychonomic Bulletin & Review*, 1–12.

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. (2003). Dissociating explicit and procedual-learning based systems of perceptual category learning. *Behavioral Processes*, *66*, 309–332.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147–161.

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

Ashby, F. G., Maddox, W. T., Glass, B. D., O'Brien, J. B., & Filoteo, J. V. (2010). Category label and response location shifts in category learning. *Psychological Research Psychologische Forschung*, *74*, 219–236.

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.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363–378.

Bamber, D. (1979). State-trace analysis: A methods of testing simple theories of causation. *Journal of Mathematical Psychology*, *19*, 137–181.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290.

Dunn, J. C., Kalish, M. L., & Newell, B. R. (2014). State-Trace Analysis can be an appropriate tool for assessing the number of cognitive systems: A reply to Ashby (2014). *Psychonomic Bulletin & Review*. (Advanced online publication) doi: 10.3758/s13423-014-0637-y

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 840–859.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211.

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science*, *21*(3), 415–423.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513.

Haslam, C., Hodder, K. I., & Yates, P. J. (2011). Errorless learning and spaced retrieval: How do these methods fare in healthy and clinical populations? *Journal of Clinical and Experimental Neuropsychology*, *33*(4), 432–447.

Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford: Oxford University Press.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear Theory, Dimensional Theory, and the Face-Inversion Effect. *Psychological Review*, *111*(4), 835–863.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*(4), 582–591.

Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning–based system in perceptual category learning. *Psychonomic Bulletin & Review*, *11*(5), 945–952.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category Number Impacts Rule-Based but Not Information-Integration Category Learning: Further Evidence for Dissociable Category-Learning Systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 227–245.

Maddox, W. T., Filoteo, J. V., & Lauritzen, J. S. (2007). Within-category discontinuity interacts with verbal rule complexity in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 197–218.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 100–107.

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, *108*(2), 578–589.

Matsuki, K. (2014). grt: General recognition theory. r package version 0.2. Retrieved from `http://CRAN.R-project.org/package=grt`

Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 676–692.

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of

COVIS. *European Journal of Neuroscience*, *34*(8), 1326–1336.

Milton, F., & Wills, A. J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 407–415.

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*(5), 563–581.

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). *6 - Systems of Category Learning: Fact or Fantasy?* (1st ed., Vol. 54). Elsevier Inc.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*, *14*(3), 505–511.

Newell, B. R., Moore, C. P., Wills, A. J., & Milton, F. (2013). Reinstating the frontal lobes? Having more time to think improves implicit perceptual categorization: A comment on Filoteo, Lauritzen, and Maddox (2010). *Psychological Science*, *24*(3), 386–389.

Nosofsky, R. M. (1986). Attention, similarity, and the identificationcategorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Potts, R., & Shanks, D. R. (2013). The Benefit of Generating Errors During Learning. *Journal of Experimental Psychology: General*.

Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, *18*(2), 414–421.

Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred Feedback Sharply Dissociates Implicit and Explicit Category Learning. *Psychological Science*, *25*(2), 447–457.

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.

Stanton, R. D., & Nosofsky, R. M. (2013). Category number impacts rule-based and information-integration category learning: A reassessment of evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1174–1191.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*.

Wills, A. J., Lea, S. E. G., Leaver, L. A., Osthaus, B., Ryan, C. M. E., Suret, M. B., ... Millar, L. (2009). A comparative analysis of the categorization of multidimensional stimuli: I. Unidimensional classification does not necessarily imply analytic processing; evidence from pigeons (Columba livia), squirrels (Sciurus carolinensis), and humans (Homo sapiens). *Journal of Comparative Psychology*, *123*(4), 391–405.

Wills, A. J., Milton, F., Longmore, C. A., Hester, S., & Robinson, J. (2013). Is overall similarity classification less effortful than single-dimension classification? *The Quarterly Journal of Experimental Psychology*, *66*(2), 299–318.

Yeates, F., Wills, A. J., Jones, F., & McLaren, I. (2012). State-trace analysis of sequence learning by recurrent networks. In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1–6).

Yeates, F., Wills, A. J., Jones, F. W., & McLaren, I. P. L. (in press). State trace analysis: Dissociable processes in a connectionist network? *Cognitive Science*.

Zaki, S. R., & Kleinschmidt, D. F. (2013). Procedural memory effects in categorization: Evidence for multiple systems or task complexity? *Memory & Cognition*, 1–17.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398.

	Strategies				
	CJ	UDO	UDL	GLC	RND
RB-FB	0.35	0.3	0.15	0.15	0.05
RB-Obs	0.4	0.2	0.1	0.2	0.1
RB Overall	0.375	0.25	0.125	0.175	0.075
II-FB	0.15	0.45	0.1	0.3	0
II-Obs	0.3	0.3	0.05	0.35	0
II Overall	0.225	0.375	0.075	0.325	0

Table 1. The proportion of participants in each condition according to the model-based analysis in Experiment 1. Conditions: RB-FB=Rule-based/feedback, RB-Obs=Rule-based/observation, II-FB=Information-integration/feedback, II-Obs=Information-integration/observation condition. Strategies: CJ=conjunction, 2D=generic two-dimensional strategy, UDO=unidimensional strategy based on orientation, UDL=unidimensional strategy based on length, GLC=general linear classifier.

	Strategies					
	CJ	2D	UD	II/OS	Other	None
RB-FB	0.45	0.05	0.1	0	0.1	0.2
RB-Obs	0.2	0.15	0.2	0	0.2	0.25
RB Overall	0.325	0.1	0.15	0	0.15	0.225
II-FB	0.4	0.25	0.25	0	0.15	0
II-Obs	0.2	0.15	0.35	0	0.05	0.25
II Overall	0.3	0.15	0.3	0	0.025	0.125

Table 2. The proportion of participants in each condition that reported using each strategy in Experiment 1. Conditions: RB-FB=Rule-based/feedback, RB-Obs=Rule-based/observation, II-FB=information-integration/feedback, II-Obs=Information-integration/observation condition. Strategies: CJ=conjunction, 2D=generic two-dimensional, UD=unidimensional, II/OS=information-integration or overall similarity.

Model-based strategies	Verbal strategy reports: Rule-based UD	CJ	2D	II/OS	Information-integration UD	CJ	2D	II/OS
UD	6	1	3	0	8	1	5	0
CJ	4	10	0	0	4	5	0	0
GLC	2	2	3	0	2	6	4	0

Table 3. Comparison of the models assigned to each participant in the model-based analysis with those they reported using in Experiment 1. UD=unidimensional, CJ=conjunction, GLC=general linear classifier, 2D=two-dimensional strategy, II/OS=either an information-integration or overall similarity strategy.

Category	Parameters μ_l	μ_o	σ_l	σ_o
A	100	200	20	20
B	100	100	20	20
B	200	100	20	20
B	200	200	20	20

Table 4. Parameters used to generate the initial stimulus distribution for Experiment 2. Each row describes a set of points in stimulus space generated by a bivariate normal distribution with means (μ_l, μ_o) and standard deviations σ_l and σ_o for the length and orientation dimensions respectively.

Condition	Strategies CJ	UDR	UDL	GLC	RND
Feedback	0.9	0	0	0.05	0.05
Observation	0.65	0	0.1	0.05	0.2

Table 5. Proportions of participants best described by each model according to the model-based analysis in Experiment 2. CJ=conjunction, UDO=unidimensional based on orientation, UDL=unidimensional based on length, GLC=general linear classifier, RND=random.

	Strategies CJ	2D	UD	II/OS	Other	None
RB-FB	0.70	0.20	0.05	0	0	0.05
RB-Obs	0.55	0.20	0.15	0	0	0.10
RB Overall	0.625	0.20	0.10	0	0	0.075

Table 6. The strategies reported by each participant in Experiment 2. Conditions: RB-FB=Rule-based/feedback, RB-Obs=Rule-based/observation. Strategies: CJ=conjunction, 2D=generic two-dimensional, UD=unidimensional, II/OS=information-integration or overall similarity.

	Verbal strategy reports			
Model-based	UD	CJ	2D	II/OS
UD	0	1	2	0
CJ	2	22	6	0
GLC	0	0	0	0
RND	2	2	0	0

Table 7. Comparison of the models assigned to each participant in the model-based analysis with those they reported using. UD=unidimensional, CJ=conjunction, GLC=general linear classifier, 2D=two-dimensional strategy, II/OS=either an information-integration or overall similarity strategies.

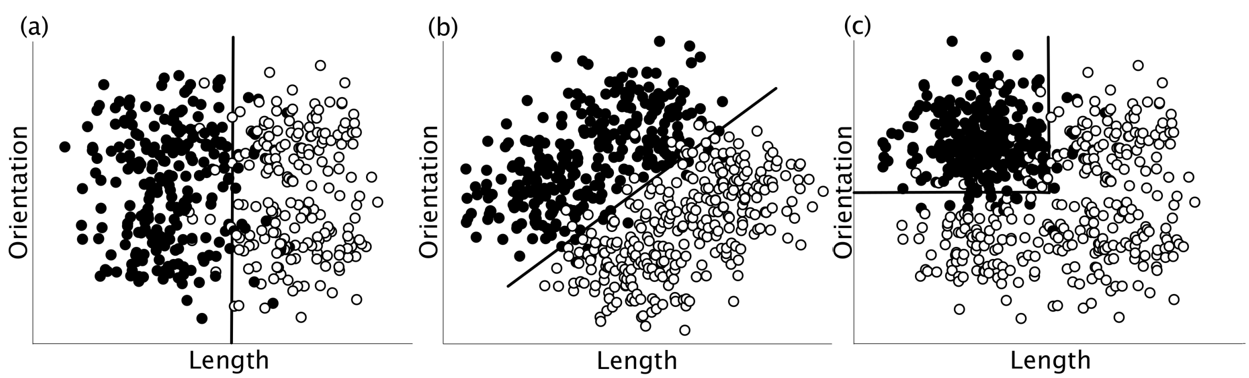

Figure 1. Stimulus space representations of (a) a unidimensional category structure, (b) a diagonal or information-integration category structure, and (c) a conjunction category structure. Filled circles represent Category A and unfilled circles represent Category B.

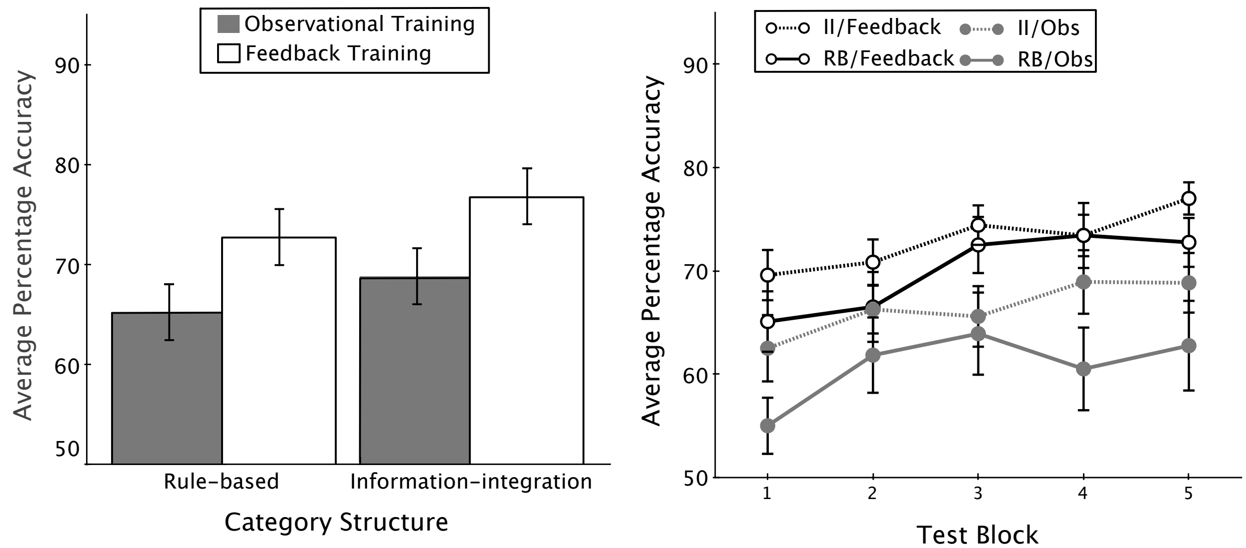

Figure 2. (a) Percentage of correct responses by condition in the final (fifth) test block. (b) The average proportion of correct responses for each block in Experiment 1. Error bars are one standard error.

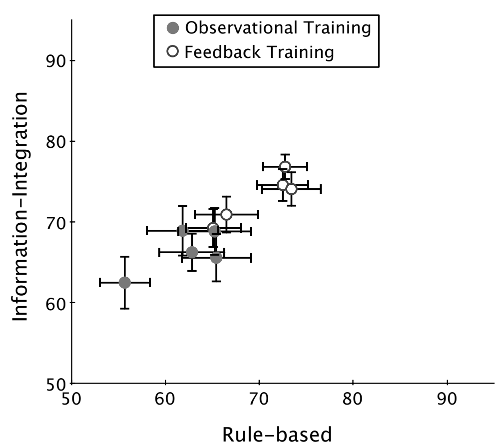

Figure 3. State-trace plot with rule-based and information-integration performance on each block on the axes. Error bars are one standard error.

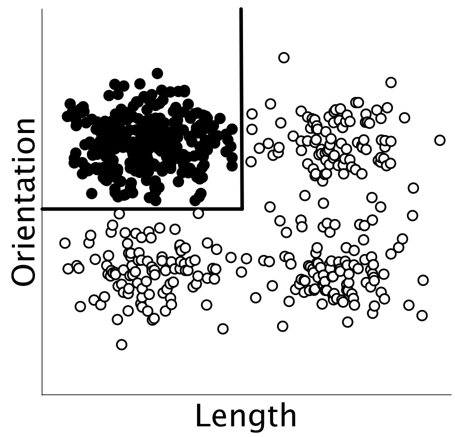

Figure 4. The conjunction category structure used in Experiment 2. Filled circles represent Category A and unfilled circles represent Category B.

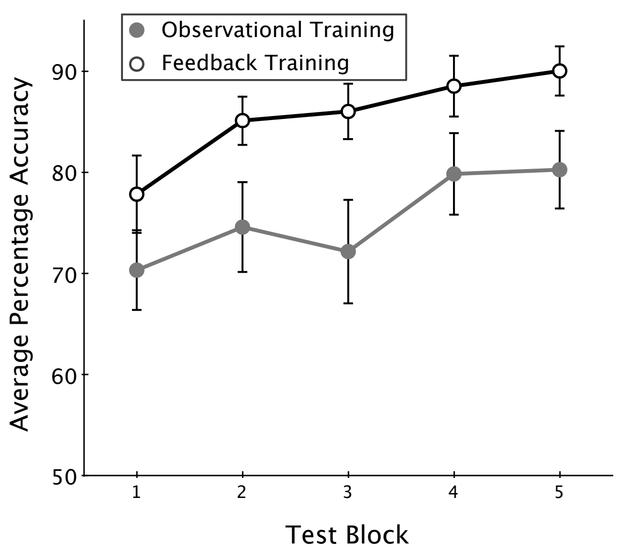

Figure 5. The average proportion of correct responses for each block in Experiment 2. Error bars are one standard error.