

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Due process in dual process:A model-recovery analysis of Smith et al. (2014)

#### **Permalink**

<https://escholarship.org/uc/item/9zj150rh>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

#### **Authors**

Edmunds, Charlotte E. R.

Milton, Fraser

Wills, Andy J.

#### **Publication Date**

2017

Peer reviewed

# Due process in dual process: A model-recovery analysis of Smith et al. (2014)

Charlotte E. R. Edmunds (ceredmunds@gmail.com)

School of Psychology, Plymouth University,  
Plymouth, PL4 8AA, UK

Fraser Milton (F.N.Milton@exeter.ac.uk)

School of Psychology, University of Exeter,  
Exeter, EX4 4QG, UK

Andy J. Wills (andy@willslab.co.uk)

School of Psychology, Plymouth University,  
Plymouth, PL4 8AA, UK

## Abstract

Considerable behavioral evidence has been cited in support of the COVIS dual-system model of category learning (Ashby & Valentin, 2016). The validity of the inferences drawn from these data critically depend on the accurate identification of participants' categorization strategies. In the COVIS literature, participants' strategies are identified using a model-based analysis inspired by General Recognition Theory (Maddox, 1999). Here, we examine the accuracy of this analysis in a model-recovery simulation. We find that participants can appear to be using implicit, procedural strategies when their responses were actually generated by explicit rule-based strategies. The implications of this for the COVIS literature are discussed.

**Keywords:** categorization; COVIS; dual-systems accounts; model-recovery; GRT

## Introduction

Categorization studies rarely examine individual differences. Rather, researchers look at group performance to draw conclusions about the likely underlying mechanisms of category learning (Kurtz, 2015). For these inferences to be valid, the participants in each group must all learn in a qualitatively similar way (Maddox, 1999). Then, relatively little information is lost by averaging. However, severe interpretative difficulties can arise if participants learn in a variety of ways, as then the average will likely not represent the behaviour of any single person (Siegler, 1987).

This issue is more than hypothetical, as there is substantial evidence, and a degree of consensus, that different participants use qualitatively different strategies in categorization tasks (e.g., Nosofsky & Zaki, 2002; Raijmakers, Dolan, & Molenaar, 2001; Wills, Inkster, & Milton, 2015). For example, some participants categorise stimuli on the basis of just one stimulus dimension (as in Figure 1B), or do so initially, even if optimum performance on the task requires using multiple stimulus dimensions (as in Figure 1B, where the participant's strategy is single-dimensional but the optimal classification strategy is diagonal).

COVIS (COmpetition between Verbal and Implicit Systems; Ashby & Valentin, 2016) is one model that aims to predict when and why participants use different strategies.

COVIS assumes that categorization is mediated by two, parallel, competing systems: an Explicit System and a Procedural System. The Explicit System is assumed to implement rule-based strategies (such as in Figure 1B). Therefore, COVIS predicts this system will optimally learn category structures that are implementations of simple rules (such as in Figure 1A). If rule-based strategies result in poor accuracy—because the category structure is not rule-based and thus difficult to verbalise—COVIS predicts the Procedural System will gain control of responding. As the Procedural System is predicted to implement a variety of strategies (including the one demonstrated in Figure 1A), it is capable of implementing the optimum strategy for information-integration category structures (the structure, but not the strategy, shown in Figure 1B).

Typical COVIS-supporting experiments look for a differential effect of an experimental manipulation (e.g. feedback timing) on rule-based and information-integration category structure learning (for a review, see Ashby & Valentin, 2016). The category structure manipulation is hoped to elicit a switch in the learning system controlling responding: because participants are learning a rule-based or information-integration category structure they will use the appropriate strategy and so be using the Explicit or Procedural System, respectively. If the experimental manipulation affects one category structure condition more than the other, the experimenter infers that it affects the accuracy of one system more than the other, thereby providing evidence for a dual-system model of category learning.

For these experiments, the presence of subsets of qualitatively-different participants can be particularly problematic. Critically, the conclusion that the experiment supports a dual-system model depends on the assumption that the participants in each category structure condition used the most appropriate system to learn those structures. In other words, that participants used the Explicit System to learn the rule-based category structure, and the Procedural System to learn the information-integration category structure. If this is not the case, any differences in overall accuracy between category structure conditions might be due to varying rates of

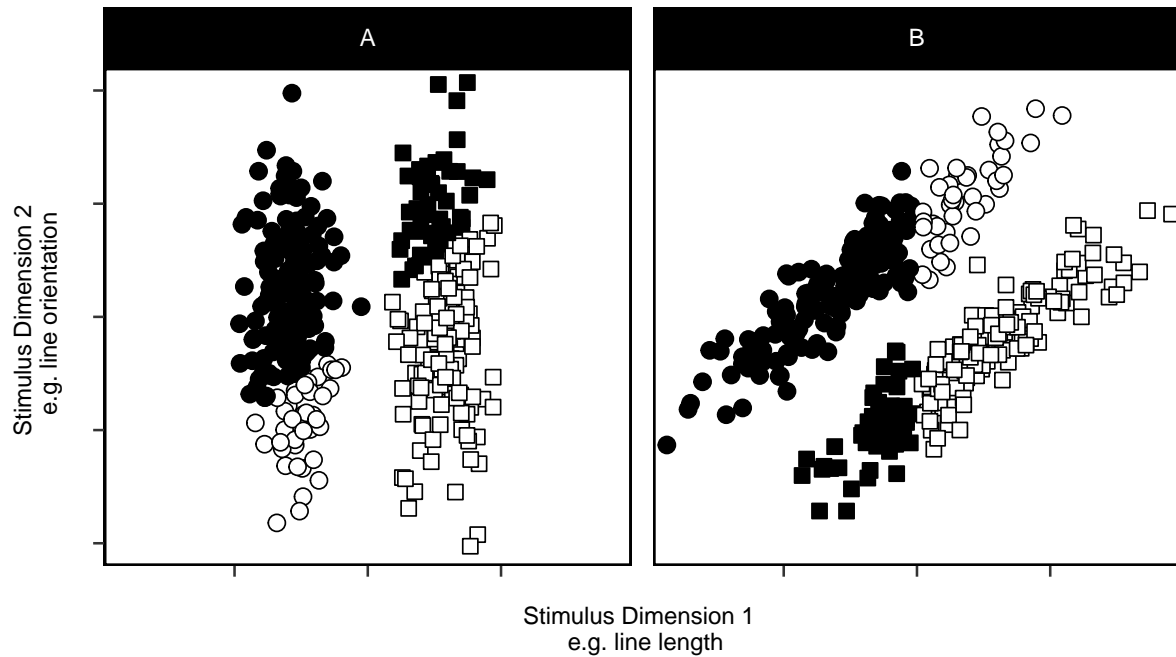


Figure 1: Two example strategies implemented on two category structures: A) a unidimensional category structure with a diagonal (GLC) strategy applied, B) an information-integration category structure with a unidimensional rule applied. Category structure: dots represent Category A, and squares Category B. Responses: Filled symbols indicate a participant responded “A”, open symbols indicate they responded “B”.

sub-optimum strategies between conditions, rather than to the existence of two category learning systems.

To illustrate this point, consider an experiment that examined the effect of deferring feedback and found it caused a reduction in performance by 10% for a particular category structure (such as in Smith et al., 2014). In the ideal case, all participants classifying a particular category structure would be using the same optimum strategy and all those in the relevant condition would be similarly affected by the manipulation; participants with deferred feedback would score 10% less than those with immediate feedback. Here, we could use standard group-accuracy analyses validly. However, if some participants were using other, sub-optimum strategies then drawing conclusions from the experiment is harder. One possibility is that the manipulation, within a given category structure, changes the relative proportions of different strategies used in each condition (feedback type). This would change average accuracy because, given a particular category structure, the highest accuracy for each strategy varies. A second possibility is the manipulation has a differential effect depending on the strategy type being used. For example, the manipulation could have had no effect on people using the optimum strategy, but could severely affect performance reliant on sub-optimum strategies (see Schnyer et al., 2009, for a similar argument).

To avoid the possibility that any dissociation in accuracy is due to the effects of sub-optimum strategies rather than

two competing systems, COVIS-supporting experiments use a strategy analysis informed by General Recognition Theory (GRT; Ashby & Gott, 1988). This analysis is used as a manipulation check to determine which strategy each participant is using. This approach (hereafter, *GRT analysis*) assumes that strategies can be modeled by a (usually linear) decision bound that passes through stimulus space (such as those in Figure 1). For each participant, a variety of strategy models are fitted to their responses. The one that best represents that participant’s pattern of responding is selected. Then, each participant’s strategy is compared to the category structure they were assigned to learn. If enough participants are found to be using the optimum strategy for the category structure they were assigned, then the category structure manipulation is assumed to have elicited a corresponding shift in category learning system. Under this assumption, any dissociations in accuracy can be validly ascribed to the existence of two systems.

Using GRT analysis as a manipulation check is logically valid if and only if GRT analysis consistently and accurately identifies participants’ strategies. In other words, GRT analysis must be able to correctly identify strategies under a variety of circumstances such as differing category structures, experimental manipulations and levels of noise. Unfortunately, recent evidence from our lab suggests that GRT analysis does not accurately recover the strategies participants use for information-integration category structures. For in-

stance, Edmunds, Milton, and Wills (2015) extended an experiment by Ashby, Maddox, and Bohil (2002) looking at feedback type by asking the participants to verbally describe their strategies. A substantial number of responders classified as using the Procedural System on the basis of GRT analysis nevertheless reported using an explicit rule-based strategy (which have been predictive of behaviour in other procedures; Wills, Milton, Longmore, Hester, & Robinson, 2013).

One possible explanation for this contradiction is that participants did not accurately report their strategies. Two pieces of evidence speak against this interpretation. First, verbal reports successfully predict participants' performance in other tasks (Lagnado, Newell, Kahan, & Shanks, 2006). Second, Carpenter, Wills, Benattayallah, and Milton (2016) found more frontal and medial temporal lobe activation for participants learning an information-integration category structure than for participants learning a rule-based structure. These brain regions are typically associated with explicit processing (Nomura et al., 2007), implying that classification of an information-integration category structure is at least as explicit in their study as classification of a rule-based structure.

A more interesting explanation for the disparity in strategies found by Edmunds et al. (2015) is that the GRT analysis is wrong. For example, because GRT analysis normally uses just the training stimuli rather than a broad range of transfer stimuli, perhaps it is biased towards the optimal strategy for each category structure? Work by Donkin et al. (2015) provides some support for this conjecture. Specifically, Donkin et al. found that including transfer stimuli from across the stimulus space reduced the proportion of participants classified as using the optimal (diagonal) strategy for an information-integration category structure.

The possibility that GRT analysis does not accurately recover the strategies participants use makes determining whether category learning is mediated by two learning systems more difficult. Consider an experiment that found that feedback delay harmed information-integration category learning but had no effect on unidimensional rule-based category learning. Furthermore, suppose that GRT analysis found that all the participants used the optimum strategy for the category structure they were presented with. If GRT analysis were accurate, we might conclude that the source of this interaction was the presence of two different systems. However, if GRT analysis was inaccurate this inference would not be the only one we could make. For example, if GRT analysis, in the information-integration conditions, falsely identified an explicit conjunction rule strategy as a diagonal (procedural) strategy, an alternative account might be that feedback delay impacts learning once participants are using sufficiently complex rules. This would be consistent with a single-system account and would potentially cast doubt on all of the COVIS-supporting studies that used this method.

However, a limitation of all work to date is that one can never be sure whether GRT analysis contains significant flaws, because one does not know which strategy participants

were actually using. When employing data from real participants, all we have are multiple forms of assessment of their strategy (GRT analysis, verbal reports, brain activations etc.), all of which provide indirect and potentially flawed information. Using one measure to assess the quality of the others includes the circularity of assuming one of the measures is correct. In the current article, we use a model-recovery approach to break out of this loop.

Model recovery involves simulating hypothetical participants' responses according to the strategy models defined by the strategy analysis. By simulating responses we circumvent many of the problems with Donkin et al. (2015) and Edmunds et al. (2015), as now we know exactly which model each (simulated) participant is using. From these hypothetical, simulated participants we can then use GRT analysis to identify the strategies from the responses to see whether GRT analysis is capable of recovering the correct generating model. This model-recovery procedure is recommended as best practice for any cognitive modeling analyses (Heathcote, Brown, & Wagenmakers, 2014) but has yet to be done for GRT analysis.

### **Simulation of Smith et al. (2014)**

Below, we use model-recovery techniques to demonstrate that current GRT analyses misidentify participants' strategies in the context of levels of performance accuracy reported in published work. Further, we demonstrate that it is possible for all participants to be using rule-based strategies but to still find a) an interaction between an experimental manipulation and category structure, and b) that the majority of participants are (incorrectly) identified by GRT analysis as using the optimum strategy for each category structure.

The experiment we chose for this demonstration is by Smith et al. (2014); a recent, representative example of empirical work within the COVIS framework (Ashby & Valentin, 2016). This experiment investigated the effect of deferring feedback on category learning. Participants were randomly assigned to learn either a rule-based or information-integration category structure (as in Figure 1) with one of two possible reinforcement schedules. In the immediate feedback condition, on each trial participants were shown a stimulus, then made their response and were immediately given corrective feedback for that trial. In the deferred feedback condition, the stimuli were shown in groups of six. The participants made responses for all six stimuli but only received corrective feedback at the end of the block. Smith et al. found that learning of the rule-based category structure was unaffected by this change in feedback timing, whereas learning of the information-integration category structure was "eliminated" (p. 454) with deferred feedback.

As well as being representative of the majority of COVIS experiments (Ashby & Valentin, 2016), the work reported in Smith et al. (2014) is interesting to simulate as it is representative of the direction that the role of GRT analysis is beginning to take in newer COVIS experiments (see also, Smith

et al., 2015). In these newer studies, the authors move away from using the GRT analysis to ensure that participants were using the optimum strategy, and therefore category learning system, in each condition. Instead, they use the GRT analysis to determine the strategies that participants use in order to discern whether deferring feedback alters the strategies participants use in “a theoretically meaningful way” (p. 452). Smith et al. (2014) conclude that deferred feedback pushed participants in the information-integration condition away from classification via the Procedural system towards classification via the Explicit system. These conclusions would of course be substantially undermined if their GRT analysis failed to correctly identify the strategies participants used.

The possibility of a misidentification of participant strategies would also open the way for an alternative, single-system, account of their results. As previously discussed, verbal report data from Edmunds et al. (2015), and neuroscience evidence from Carpenter et al. (2016), indicate that participants sometimes learn information-integration category structures using complex, verbalisable rules—despite the GRT analysis pointing towards procedural (GLC) strategies in these cases. Perhaps this is also happening in Smith et al. (2014)? Specifically, we hypothesize that the majority of participants in the immediate information-integration category structure condition of Smith et al. are using a conjunction or another two-dimensional rule-based strategy, but this is mis-identified as an implicit (GLC) strategy by Smith et al.’s GRT analysis. The possibility of this kind of mis-identification seems particularly acute in this study because those authors did not include a conjunction rule (or any other complex rule) in the set of models for their GRT analysis. Research by Donkin et al. (2015) suggests that failing to include complex rules in a GRT analysis increases the proportion of participants that are identified as procedural (GLC) responders.

## Method

To see whether it was possible that all the participants in Smith et al. (2014) were using rule-based strategies, we first generated a set of hypothetical participants. These participants’ responses were generated from unidimensional and conjunction strategy GRT generating models that best fit either the unidimensional or information-integration category structures used by Smith et al. The *unidimensional* models were straight lines that passed perpendicularly through either the *x*-axis or the *y*-axis. Stimuli that lay on one side of the line were assigned “Category A” and those on the other “Category B.” The *conjunction* models consisted of two lines perpendicular to each other that partitioned off a quarter of the space. The stimuli in that quarter were assigned “Category A” and those outside “Category B.”

We then added various levels of noise to these hypothetical participants and calculated their accuracy. Twenty participants were generated for each level of noise, category structure and generating strategy. Then we performed the GRT analysis, which included three model types: unidi-

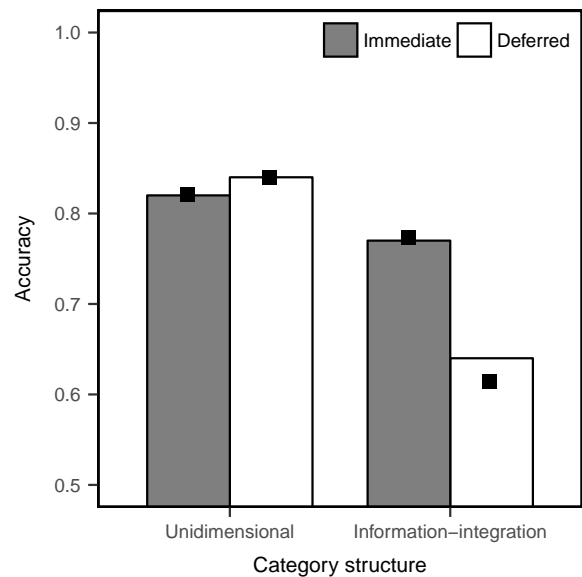


Figure 2: Simulation of Smith et al. (2014); bars are empirical data; plot points are the simulation. Smith did not report standard deviation.

dimensional, diagonal (GLC) and random models (Maddox & Ashby, 1993). Note that although some simulated participants’ responses were generated by a conjunction strategy, this strategy type was not included in the GRT analysis. This was to keep the GRT analysis as similar as possible to the one conducted by Smith et al. (2014). We then selected 21 simulated participants (i.e. the same *N* as Smith et al., 2014) for each condition such that, as far as was possible, they had a) the same average accuracy as that reported by Smith et al. (p. 451, their paper; Figure 2, current paper), b) the same number of “strong learners” (p. 541, their paper), and c) were identified by GRT analysis as using the same distribution of strategy types reported by Smith et al. (p. 452-453, their paper; Table 1, this paper).

## Results

In addition to the simulated participants having the same average accuracy (see Figure 2) and same distribution of GRT-recovered strategies (see Table 1) as the real participants in Smith et al. (2014), it was also possible to replicate Smith et al.’s statistical tests. For the simulated participants, the critical interaction between category structure and task was significant,  $F(1, 80) = 10.64, p = .002$ . Furthermore, as in Smith et al. (2014), performance in the two rule-based conditions were statistically indistinguishable,  $t(40) = 0.44, p = .663$ , as was the comparison between the unidimensional and information-integration immediate conditions,  $t(40) = 1.22, p = .228$ . Whereas, the difference between the two information-integration category structure conditions did reach significance,  $t(40) = 4.98, p < .001$

Table 1 shows that it is possible to generate the statisti-

cal pattern and strategy model results reported in Smith et al. (2014), without resorting to a second Procedural System. Instead, all the so-called implicit responders found by Smith et al. (2014) could have been using rule-based strategies that were misidentified by the GRT analysis.

Table 1: GRT analysis of simulated participants for Smith et al. (2014). Counts in **bold** are from real participants, as reported by Smith et al. (2014), and are also the simulation results (the simulation exactly reproduces the observed distribution of recovered models). Remaining counts show how the two groups of generating models used in this simulation (UD and CJ) were recovered by the GRT analysis. So, for example, of the 18 UD generating models used in the UD-Imm condition, 13 were correctly recovered as UD.

|                | Recovered strategies |           |           |          |
|----------------|----------------------|-----------|-----------|----------|
|                | UDx                  | UDy       | GLC       | RND      |
| <b>UD-Imm.</b> | <b>13</b>            | <b>1</b>  | <b>1</b>  | <b>6</b> |
| Gen. model: UD | 13                   | 1         | 1         | 3        |
| Gen. model: CJ | 0                    | 0         | 0         | 3        |
| <b>UD-Def.</b> | <b>15</b>            | <b>2</b>  | <b>0</b>  | <b>4</b> |
| Gen. model: UD | 15                   | 2         | 0         | 2        |
| Gen. model: CJ | 0                    | 0         | 0         | 2        |
| <b>II-Imm.</b> | <b>0</b>             | <b>3</b>  | <b>16</b> | <b>2</b> |
| Gen. model: UD | 0                    | 3         | 0         | 1        |
| Gen. model: CJ | 0                    | 0         | 16        | 1        |
| <b>II-Def.</b> | <b>2</b>             | <b>13</b> | <b>3</b>  | <b>3</b> |
| Gen. model: UD | 2                    | 4         | 0         | 0        |
| Gen. model: CJ | 0                    | 9         | 3         | 3        |

Strategies: UDx = Unidimensional based on the x-dimension, Udy = Unidimensional based on the y-dimension, GLC = General linear classifier, RND = Random.

## General Discussion

The influential COVIS model of category learning is supported by a great deal of behavioural data (Ashby & Valentin, 2016). Predominantly, this evidence comes from a single experimental methodology which examines the effect of a factor on rule-based and information-integration category learning. COVIS predicts that its two systems can implement different strategy types, and so each will learn one of these category structures better than the other. Critically, the validity of the inferences from this paradigm hangs on correctly identifying the strategy each individual used to complete the learning task. This is because the experiments investigating COVIS cannot directly control which system participants use to respond. Instead, they manipulate the category structures and hope that this encourages participants to use the optimum system, and thus the correct strategy, for that category structure. Of course, participants may continue to use the sub-optimum system for a particular category structure. Thus, identifying the strategies participants use is crucial: if the participants are using the correct strategy for that category structure, then the experimenters assume that they must also be using the cor-

rect learning system for that structure. Then, any differential effects of a manipulation on each category structure can be attributed to the existence of two systems of category learning, not differing numbers of sub-optimal responders.

Despite its importance for the COVIS model, there is experimental (Edmunds et al., 2015) and modeling (Donkin et al., 2015) evidence to suggest that GRT analysis may be biased towards concluding that participants were using the optimum strategy for the category structure. To explore this possibility, we simulated an experiment by Smith et al. (2014) and showed that it was possible to reproduce their means, inferential statistics and strategy analysis using only participants who used rule-based strategies. Simulated participants classified the information-integration category structure using a conjunction rule, but were recovered by the strategy analysis as using a diagonal (GLC) strategy. This raises the possibility that participants in Smith et al. were, correspondingly, using rule-based strategies in classifying the information-integration category structure. In other words, Smith et al. cannot be construed as clear evidence for dual-system accounts of category learning, as a single-system (rule-based) account also fits all the data (accuracy and GRT analysis) they presented.

## Implications for the COVIS model

The reported simulation demonstrates an inferential weakness in experiments argued to support COVIS: GRT analysis is not accurate enough to act as a manipulation check. It cannot determine whether manipulating the category structure successfully elicited a corresponding switch in the categorisation system underlying participants' responses. Consequently, it is difficult to judge whether a particular COVIS-supporting dissociation is due to the existence of two distinct learning systems, or rather due to participants using different explicit strategies to learn each category structure. This increases uncertainty over conclusions of a swathe of COVIS-supporting studies that rely on comparing rule-based and information-integration category structures (see Ashby & Valentin, 2016, for a partial list).

In relation to the experimental work by Edmunds et al. (2015), and Edmunds, Wills, and Milton (2016), this simulation also strengthens the evidence that participants can correctly report their categorisation strategies. In those experiments, participants learning information-integration category structures consistently reported using complex, rule-based strategies. In contrast, the GRT analysis identified these participants as using the correct (i.e. diagonal) strategy. In the above simulation, it was shown that participants using a conjunction rule were likely to be misidentified in GRT analysis as using a diagonal (GLC) strategy. Therefore, it seems plausible that all participants learn information-integration category structures explicitly, using rule-based approaches, but GRT analysis misidentifies some of these as using an implicit (GLC) strategy.

## Conclusions

The simulation reported above indicates that drawing conclusions from GRT analysis is risky. This has a knock on effect on the COVIS-supporting studies that rely on this analysis as a manipulation check. More investigations need to be done to understand which strategies participants use and how they are affected by the category structure being learned before we can be sure that experimental dissociations in this literature support a dual-system model of categorization. In other words, we advocate closer attention to due process in the evaluation of dual-system (and single-system) models.

## References

- 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. doi: 10.1037/0278-7393.14.1.33
- 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. doi: 10.3758/BF03196423
- Ashby, F. G., & Valentin, V. V. (2016). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (2nd ed., pp. 547–572). New York, NY: Elsevier.
- Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016). A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping*, *37*, 3557–3574. doi: 10.1002/hbm.23259
- Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., Nosofsky, R. M., Donkin, C., ... Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 933–948. doi: 10.1037/xlm0000083
- Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *The Quarterly Journal of Experimental Psychology*, *68*(2), 1203–1222. doi: 10.1080/17470218.2014.978875
- Edmunds, C. E. R., Wills, A. J., & Milton, F. N. (2016). Memory for exemplars in category learning. In A. Papfragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2243–2248). Austin, TX: Cognitive Science Society.
- Heathcote, A., Brown, S. D., & Wagenmakers, E.-j. (2014). *An introduction to good practices in cognitive modeling*. New York, NY: Springer.
- Kurtz, K. J. (2015). Human category learning: Toward a broader explanatory account. In *Psychology of learning and motivation* (Vol. 63, pp. 77–114). Academic Press. doi: 10.1016/bs.plm.2015.03.001
- Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, *135*(2), 162–183. doi: 10.1037/0096-3445.135.2.162
- Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics*, *61*, 354–74. doi: 10.3758/BF03206893
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49–70. doi: 10.3758/BF03211715
- Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, a. D., Gitelman, D. R., Parrish, T. B., ... Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*(1), 37–43. doi: 10.1093/cercor/bhj122
- Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924–940. doi: 10.1037/0278-7393.28.5.924
- Rajmakers, M. E. J., Dolan, C. V., & Molenaar, P. C. M. (2001). Finite mixture distribution models of simple discrimination learning. *Memory & Cognition*, *29*(5), 659–677. doi: 10.3758/BF03200469
- Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. *Neuropsychologia*, *47*(13), 2995–3006. doi: 10.1016/j.neuropsychologia.2009.07.011
- Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*(3), 250–264. doi: 10.1037/0096-3445.116.3.250
- 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), 447–57. doi: 10.1177/0956797613509112
- Smith, J. D., Zakrzewski, A. C., Herberger, E. R., Boomer, J., Roeder, J. L., Ashby, F. G., & Church, B. A. (2015). The time course of explicit and implicit categorization. *Attention, Perception, & Psychophysics*, *77*(7), 2476–2490. doi: 10.3758/s13414-015-0933-2
- Wills, A. J., Inkster, A. B., & Milton, F. (2015). Combination or differentiation? Two theories of processing order in classification. *Cognitive Psychology*, *80*, 1–33. doi: 10.1016/j.cogpsych.2015.04.002
- 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. doi: 10.1080/17470218.2012.708349