Australian Journal of Psychology

Australian Journal of Psychology 2012; **64**: 46–51
doi:10.1111/j.1742-9536.2011.00035.x

Levels of explanation in category learning

Ben R. Newell

School of Psychology, University of New South Wales, Sydney, NSW, Australia

Abstract

Multiple-system accounts of category learning are now entrenched in the cognitive neuroscience literature. This entrenchment sometimes seems impervious to behavioural evidence that contradicts or questions key assumptions of multiple-systems accounts. In this brief article, I consider relevant sources of evidence (behavioural, neuropsychological, neuro-imaging) and argue that the evidence from all sources is not as clear-cut as many multiple-system theorists often claim. More importantly, the review emphasises that one needs to be sensitive to the desired level of explanation of category learning (psychological, biological, mathematical) when considering the relevance of different types of data and the adequacy of proposed accounts.

Key words: category learning, cognitive neuroscience, single-system, multiple-systems

Imagine, just for the sake of argument that you had submitted an article for publication that reported data which, in your opinion, challenged the dominant perspective in a particular area. Imagine that the area was category learning and that the 'challenge' was to the much vaunted view that multiple systems subserve this cognitive ability. Finally, imagine that one of the reviews you received contained something along these lines:

> This manuscript completely ignores the neuroscience evidence in support of multiple systems. In short, the neuroscience evidence in support of multiple systems of category learning is overwhelming. I think that nearly every cognitive neuroscientist takes the existence of multiple systems as fact, and those who are versed in the category learning literature think the same way. The verdict is already in, the existence of multiple systems is decided and the field should focus on elucidating the nature of these systems by understanding the ways in which they cooperate and compete.

The review is of course a caricature which serves to illustrate the extremity of the position I want to examine; but I'd hazard a guess that the opinion is not so very far from that of some of my colleagues. What are the implications of this view? Should those of us who have reservations about multiple-system interpretations pack up and go home? Are we wasting our time?

Correspondence: Ben R. Newell, Associate Professor, School of Psychology, University of New South Wales, Sydney NSW 2052, Australia. Email: ben.newell@unsw.edu.au

Received 5 April 2011. Accepted for publication 23 June 2011.

The aim of this brief article is to highlight some of the reasons why, in my opinion, the issue of the existence of multiple systems in category learning (and indeed many other domains) is not as cut-and-dried as our hypothetical reviewer might have us believe.

The article proceeds as follows: Section I examines what is meant by a 'system'; Section II reviews very briefly some of the behavioural evidence that speaks 'for' and 'against' the existence of multiple systems in category learning; Section III highlights some relevant neural evidence, and Section IV asks how we might reconcile perspectives at different levels of explanation. The coverage of these issues is necessarily brief: The interested reader can find an extended treatment of many of the topics in Newell, Dunn, and Kalish (2011).

WHAT IS A SYSTEM?

This question turns out to be a very tricky one to answer. For example, how does one reconcile the fact that the brain is part of the single central nervous system, but that different parts of the brain are specialised for particular functions (e.g., the visual *system*)? One approach is to consider the different 'levels' (of explanation) at which one might want to describe something as an independent system. Ashby and Ell (2002) suggest a hierarchy of criteria from mathematical to psychological to neurobiological. At the mathematical level, a system is defined as the mapping of input variables to output variables via a set of parameters. For two systems to be independent, there must be some set of parameter values that make unique predictions for each system (i.e., the systems must not be identical nor fully nested).

The psychological and neurobiological criteria in Ashby and Ell's framework map, roughly, on to the structural and

brain levels, respectively in Keren and Schul's (2009) recent analysis. The structural/psychological level refers to 'the kind of processing they do [and] the type of representation they use' (Keren & Schul, p.536), or the psychological processes that are required by a task (Ashby & Ell, 2002). For something to count as a separate system at this level there must be clear differences in the nature of the process and representation being utilised. At the brain/neurobiological level, the simple (simplistic?) idea is to show that separate systems are mediated by unique structures or pathways in the brain (e.g., Poldrack, 2010).

A final level, considered by Keren and Schul (2009) (but not explicitly by Ashby & Ell), is the *functional* one (cf. Marr, 1982). A useful notion when considering the functional level is that of *functional incompatibility* (Sherry & Schacter, 1987). This criterion is satisfied when a system has evolved because the function that it serves cannot be performed by any existing system within the organism. An example, provided by Sherry and Schacter (1987) is the dual visual system that insects have developed to control orientation in flight. Ocelli eyes (located frontally) can respond rapidly to changes in average light intensity but have poor spatial resolution. In contrast, compound eyes (located laterally) have fine spatial resolution but respond slowly. Thus, the functions that one 'visual-system' serves is incompatible with the other—but both help stabilise insects against roll and pitch in flight.

I think this notion of function is particularly useful when it comes to category learning. A question one might ask is the degree to which the functions of the two proposed systems are incompatible with each other. For example, would we have evolved functionally independent systems for making 'complex' and 'simple' discriminations (e.g., discriminating X-rays that show/do not show tumours vs discriminating circles from squares—see Ashby & Maddox, 2005)?

BEHAVIOURAL EVIDENCE

The majority of multiple (mostly dual) system models in the category learning literature are described in terms of their psychological properties. One system, variously termed explicit, declarative, verbal or rule-based relies on working memory, hypothesis testing, and the application of simple rules (Ashby & Maddox, 2005; Minda & Miles, 2010). The other, described as implicit, procedural, non-verbal or similarity-based does not involve working memory or attention and learns associations between (motor)responses and category labels (Ashby & Maddox, 2005; Minda & Miles, 2010).

This characterisation leads naturally to a number of predictions about the impact of various task factors on category structures claimed to be learned by one or other system. For example, several studies have argued that tasks solved via simple rules are affected by the addition of cognitive load tasks during learning, whereas more complex tasks, amenable to learning via the implicit system, are not (Foerde, Knowlton & Poldrack, 2006; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). However, in each of these cases, there is reason to question the selective effects of load (Newell, Dunn, & Kalish, 2010; Newell, Lagnado, & Shanks, 2007; Nosofsky & Kruschke, 2002; Nosofsky, Stanton, & Zaki, 2005) suggesting that the proposed dissociation is not clear-cut.

Likewise, experiments that have emphasised the facilitative effect of working memory for rule-based tasks and its detrimental effect for non-rule-based ones (e.g., DeCaro, Thomas, & Beilock, 2008) have been re-evaluated (Tharp & Pickering, 2009) or have not been replicated (e.g., Lewandowsky, 2011).

Moreover, although the product of learning from the 'implicit' system is often assumed to be unavailable to awareness or impossible to verbalise (Ashby, Alfonse-Reese, Turken, & Waldron, 1998; Knowlton, Squire, & Gluck, 1994; Minda & Miles, 2010), several studies show levels of awareness in 'signature' implicit tasks that are sufficient to account for participants' performance (e.g., Lagnado, Newell, Kahan, & Shanks, 2006; Newell et al., 2007; Speekenbrink & Shanks, 2010). Such findings leave little room for the contribution of an implicit system.

Quite apart from the 'to-and-fro' of behavioural evidence variously interpreted as favouring dual or single system accounts, there are some deeper concerns about the whole enterprise of relying on functional dissociations for determining the dimensionality of data (Newell & Dunn, 2008). Put simply, finding that variable A (e.g., cognitive load) has a detrimental effect on task X but no effect on task Y tells you precisely nothing about whether task X and task Y are subserved by qualitatively different systems. Even if task X is detrimentally affected while task Y is simultaneously facilitated—a 'double dissociation'—this still does not allow—logically—conclusions to be drawn regarding the number of systems (or independent processes) producing the data (Dunn & Kirsner, 1988).

It is beyond the scope of this brief article to discuss the details of why the dissociation logic is flawed. Perhaps the simplest way to intuit the problem is to realise that dissociation logic requires one to be sure that a given variable has *absolutely no effect* on a task—and this is impossible (Dunn, 2003). We have advocated an alternative approach based on *state-trace analysis* (Bamber, 1979) which allows one to examine the dimensionality of data without relying on dissociation logic. Our applications of this approach to category learning tasks has, so far, led us to conclude that a single system is sufficient to account for performance (see Newell &

Dunn, 2008; Newell et al., 2010, 2011 for details of the analysis and results)[1].

In summary, at the behavioural level, the status of multiple-system explanations is not clear: Can the neuroscience evidence clarify the picture?

NEUROSCIENCE EVIDENCE

As our caricatured reviewer reminds us the *real action* in determining the nature of category learning lies at the neural level: What does the overwhelming neuroscience evidence look like? Two classes of evidence can be distinguished—(1) neuropsychological studies showing that different patient groups (e.g., schizophrenics, Parkinson's Disease) are impaired/not impaired on implicit/explicit versions of category learning tasks (e.g., Knowlton, Mangels, & Squire, 1996; Price, Filoteo, & Maddox, 2009); and (2) neuro-imaging data of 'normal' participants suggesting the involvement of separable systems in implicit and explicit tasks (e.g., Nomura & Reber, 2008; Reber, Gitelman, Parrish, & Mesulam 2003).

In line with the behavioural evidence, although the neural data is often portrayed as overwhelming (e.g., Poldrack & Foerde, 2008) there is reason to be cautious with both these sources of evidence.

With regard to the neuropsychological evidence, several studies suggest that adopting a one-to-one mapping of deficits in category learning with functionally discrete learning and memory systems in the brain is naive (cf. Palmeri & Flanery, 2002). Purportedly diagnostic dissociations in a variety of patient populations have been challenged by several authors (e.g., Kinder & Shanks, 2003; Nosofsky & Zaki, 1998). The most enduring alternative explanation is that all participants, whether brain function is compromised or not, adopt an explicit hypothesis testing approach to learning tasks often characterised as 'classic' implicit tasks (e.g., the 'weather prediction task'—see Speekenbrink, Channon, & Shanks, 2008; Speekenbrink, Lagnado, Wilkinson, Jahanshahi, & Shanks, 2010).

Reconsideration of the neuropsychological literature also suggests that drawing inferences from the aetiology of certain conditions (e.g., degeneration in dopamine containing cells in Parkinson's disease sufferers) to observed deficits on category learning tasks is naive for at least two reasons. First, tasks that are often grouped together as generic 'explicit' and 'implicit' tasks may be affected differently by manipulations that are predicted to have similar effects (Price, 2005). Second, there is clear evidence that patients on and off medication behave very differently and in ways that are often at odds with proposed dissociations (e.g., Jahanshahi, Wilkinson, Gahir, Dharminda & Lagnado, 2010; Swainson et al., 2006; Wilkinson, Lagnado, Quallo, & Jahanshahi, 2008).

It is also the case that neuropsychology relies—*par excellence*—on dissociation logic (Dunn, 2003); but when one applies alternative approaches—such as state-trace analysis, data patterns that appear under dissociation logic to support differential involvement of systems, no longer necessarily lead to such a conclusion. For example, Newell et al. (2011) used the principles of state-trace analysis to examine category learning data from Parkinson's Disease patients, Huntington Disease's patients, and normal controls (data originally collected by Filoteo, Maddox, & Davis, (2001) and Filoteo, Maddox, Salmon, & Song 2005). This re-examination demonstrated that, contrary to the original interpretation, the data did not provide strong evidence against a single-system interpretation (see fig. 7, p.195 in Newell et al., 2011).

The remit of the neuro-imaging studies is to demonstrate that region X subserves an 'implicit' task while region Y is involved in an 'explicit' task. Traditionally, studies have indicated that the X is the basal ganglia system (primarily the striatum), and Y is the medial temporal lobes, in particular the hippocampus (e.g., Squire, 1992). But how specific and pure are these mappings between regions and tasks?

Evidence from category learning tasks increasingly shows the involvement of *both* systems in *both* kinds of tasks (e.g., Dickerson, Li, & Delgado, 2011; Foerde et al., 2006; Poldrack et al., 2001). Thus, the debate in the neuro-imaging literature has shifted to questions of how the two systems interact: Do they compete with one another (e.g., Poldrack et al., 2001) or operate in parallel (e.g., Dickerson et al., 2011)? These are important questions, but they can be answered in different ways depending on the *level of explanation* that one subscribes to.

For example, Dickerson et al. (2011) conclude from a neuro-imaging study that used a variant of a commonly used category learning task that 'these results suggest that distinct human memory systems operate in parallel during probabilistic [category] learning, and may act synergistically . . . to jointly contribute to learning and decision-making' (p.266). However, at the level of observed behaviour (i.e., accuracy in category learning) Dickerson et al. found that a task difficulty variable had statistically identical effects on purportedly explicit and implicit versions of their task. The functional neuro-imaging identified regions of interest (ROIs) in the striatum and the hippocampus that showed an effect of task difficulty and found that these ROIs displayed similar patterns of BOLD responses *irrespective* of task type (i.e., 'implicit vs explicit'). Furthermore, these regions were *functionally* correlated. Thus, the evidence that the systems are distinct appears to come purely from their distinct anatomical location in the brain (i.e., one interested region is in the basal ganglia, and one is in the medial temporal lobe) and *not* from their *functional* contribution to category learning. Such a conclusion begs the question of when *two*

systems that operate in parallel are functionally correlated and contribute 'synergistically' to the same patterns of observable behaviour become *one* system.

Consistent with the behavioural data, neuroscience evidence does not appear to point unequivocally to independent, qualitatively, and/or functionally different systems. And yet, the multiple-systems interpretation remains pervasive. Is there any way to reconcile multiple and single-system perspectives?

RECONCILING DIFFERENT LEVELS OF EXPLANATION

In a recent set of articles, Poldrack (2010) and Shimamura (2010) discuss the 'what' and 'where' strategies that researchers can adopt in investigating a phenomenon. For category learning, the 'what' strategy might be a general question like 'What is category learning?' As Shimamura notes, to answer such a question, one needs to describe mental events in terms of their psychological processes—i.e., the *structural* level of explanation. Such explanations can be entirely agnostic about underlying brain processes—they can explain *our capacity* to categorise without being concerned about how this capacity is achieved at the neural level (cf. Bennett & Hacker, 2003; Trigg & Kalish, 2011). Importantly, the quality of such explanations can be assessed (independently of neural evidence) by their ability to explain behavioural data or to lead to computational models that can successfully implement verbal theories and predictions (e.g., Lewandowsky & Farrell, 2010). Thus, explanations of the 'what' question are by no means diminished (indeed they are unaffected) by the absence of neural specification because this is not what they aim to explain.

The question then is whether the 'where' strategy—questions that attempt to define the neural correlates of a psychological phenomenon—aid our understanding. Does knowing *where* category learning occurs improve our understanding of *what* category learning is? This is where the debate heats up. Our caricatured reviewer clearly thinks the answer is yes: More than that, he/she thinks that 'where' evidence should take precedence over apparently contradictory 'what' evidence. But what is the basis for such a hierarchy of evidential value? A response might be that brain activity must be the basis for psychological processes, and thus if one's theory 'gets the brain bit wrong', then it must also have the psychological explanation wrong. Such an argument is, of course, flawed however not least because given our current understanding of brain science, it is highly unlikely that a neural explanation is entailed by a psychological one (Newell et al., 2011).

Shimamura (2010) suggests that a possible way forward is for cognitive neuroscientists to ask the 'How' questions—e.g., *how* does the striatum contribute to category learning? He argues that doing so will necessitate considering broader neural circuits and 'whole brain interactions' (p. 774). This is a useful recommendation and would, as Shimamura notes, allay some of the 'narrow localisation' critiques often aimed at cognitive neuroscience. Following such advice, however, will require extremely careful experimentation and clear thinking about exactly what functions we are measuring at the various levels of analysis.

A good example of the erroneous conclusions that can be drawn when there is confusion about the relation between brain-level and structural (psychological)-level explanations comes from a recent neuro-imaging study by Gureckis, James, and Nosofsky (2011). Gureckis et al. (2011) followed up a study by Reber et al. (2003) in which it was claimed that participants learning a classification task under explicit (intentional) or implicit (incidental) conditions relied on separable implicit (striatum) and explicit (hippocampus) neural systems. Although Gureckis et al. observed similar differences in neural activity in their participants; the authors were able to demonstrate that the differential activity was more readily interpreted as due to differences in the specific stimulus-encoding instructions given to participants in the implicit and explicit conditions than to the recruitment of distinct neural systems. In other words, careful thinking about the structural level of explanation (the psychological processes engaged in mediating learning) led to a re-evaluation of the brain-level explanation. This re-evaluation emphasises the need for caution when drawing strong conclusions on the basis of evidence from neuro-imaging, especially given that the Reber et al. result had previously been cited as one of the strongest pieces of evidence for the involvement of separable systems (e.g., Poldrack & Foerde, 2008).

CONCLUSION

Some researchers have accused those who have reservations about multiple-system interpretations of category learning as 'protecting [their] central assumptions and attacking results from the MMS [multiple memory systems] approach rather than inspiring novel findings' (Poldrack & Foerde, 2008, p.202). Such claims are inaccurate and unhelpful—in much the same way as the blinkered caricature reviewer who admits no room for argument. If we are to understand category learning, we need first to identify the level of explanation that we desire—what, where, how, functional—and acknowledge that different types of evidence pertain to different levels. To suggest that one type of evidence is privileged or precludes consideration of other types will only lead to continued acrimony of the kind exemplified in the Poldrack and Foerde, 2008 quote. Once, we are clear with each other

about what it is that we are trying to explain, I am optimistic that we will be able to employ the appropriate methodological and conceptual analyses to reach a clearer understanding of how *we* (and our brains) categorise our world.

NOTE

1. It should be noted that if state-trace analysis finds evidence for more than one latent variable or underlying parameter varying across conditions in a dataset, this is not necessarily inconsistent with 'single-system' models. State-trace analysis is agnostic with respect to what the dimensionality of the data represents—additional dimensions could be equally consistent with multiple-system models or single-system models that might predict variation in more than one parameter across conditions.

ACKNOWLEDGEMENTS

The support of the Australian Research Council (Grant DP: 0877510 awarded to the author, John Dunn and Mike Kalish) is gratefully acknowledged. I thank John Dunn and Mike Kalish for extensive discussion of many of the issues outlined in the article and Rob Nosofsky for very helpful comments on an earlier draft of this article.

REFERENCES

Ashby, F. G., & Ell, S. W. (2002). Single versus multiple systems of learning and memory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology Vol. 4: Methodology in experimental psychology* (3rd ed., pp. 655–691). Hoboken, NJ: John Wiley & Sons Inc.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178. doi: 10.1146/annurev.psych.56.091103.070217

Ashby, F. G., Alfonse-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481. doi: 10.1037/0033-295X.105.3.442

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

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell Publishing.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*, 284–294.

Dickerson, K. C., Li, J., & Delgado, M. R. (2011). Parallel contributions of distinct human memory systems to probabilistic learning. *Neuroimage*, *55*, 266–276.

Dunn, J. C. (2003). The elusive dissociation. *Cortex*, *39*, 177–179.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological. Review*, *95*, 91–101.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001). A possible role of the striatum in linear and nonlinear categorization rule learning: Evidence from patients with Huntington's disease. *Behavioral Neuroscience*, *115*, 786–798.

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology*, *19*, 212–222.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 11778–11783.

Gureckis, T. M., James, T. W., & Nosofsky, R. M. (2011). Reevaluating dissociations between implicit and explicit category learning: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *23*, 1697–1709.

Jahanshahi, M., Wilkinson, L., Gahir, H., Dharminda, A., & Lagnado, D. A. (2010). Medication impairs probabilistic classification learning in Parkinson's disease. *Neuropsychologia*, *48*, 1096–1103. doi: 10.1016/j.neuropsychologia.2009.12.010

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533–550. doi: 10.1111/j.1745-6924.2009.01164.x

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review*, *110*, 728–744. doi: 10.1037/0033-295X.110.4.728

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402. doi: 10.1126/science.273.5280.1399

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*, 106–120.

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*, 162–183. doi: 10.1037/0096-3445.135.2.162

Lewandowsky, S. (2011). Working memory capacity and categorisation: Individual differences and modelling. *Journal of Experimental Psychology, Learning, Memory & Cognition*, *37*, 720–738.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.

Marr, D. (1982). *Vision*. San Francisco: H. Freeman and Co.

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. In B. H. Ross (Ed.), *The Psychology of learning and motivation: Advances in research and theory (Vol. 52)* (pp. 117–162). San Diego: Academic Press.

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*, 285–290. doi: 10.1016/j.tics.2008.04.009

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*, 563–581. doi: 10.3758/MC

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? *The Psychology of Learning & Motivation*, *54*, 167–215.

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

Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews*, *32*, 279–291. doi: 10.1016/j.neubiorev.2007.07.006

Nosofsky, R. A., & Kruschke, J. K. (2002). Single system models and interference in category learning: Commentary on Waldron & Ashby (2001). *Psychonomic Bulletin & Review*, *9*, 169–174.

Nosofsky, R. A., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, *33*, 1256–1271.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255. doi: 10.1111/1467-9280.00051

Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research theory, (Vol. 41)* (pp. 141–189). San Diego: Academic Press.

Poldrack, R. A. (2010). Mapping mental function to brain structure: How can cognitive neuroimaging succeed? *Perspectives in Psychological Science*, *5*, 753–761.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews*, *32*, 197–205. doi: 10.1016/j.neurobiorev.2007.07.007

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550. doi: 10.1038/35107080

Price, A. L. (2005). Cortico-striatal contributions to category learning: Dissociating the verbal and implicit systems. *Behavioral Neuroscience*, *119*, 1438–1447. doi: 10.1037/0735-7044.119.6.1438

Price, A., Filoteo, J. V., & Maddox, W. T. (2009). Rule-based category learning in patients with Parkinson's disease. *Neuropsychologia*, *47*, 1213–1226. doi: 10.1016/j.neuropsychologia.2009.01.031

Reber, P., Gitelman, D., Parrish, T., & Mesulam, M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, *15*, 574–583.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, *94*, 439–454. doi: 10.1037/0033-295X.94.4.439

Shimamura, A. P. (2010). Bridging psychological and biological science: The good, the bad, and ugly. *Perspectives in Psychological Science*, *5*, 772–775.

Speekenbrink, M., Channon, S., & Shanks, D. R. (2008). Learning strategies in amnesia. *Neuroscience and Biobehavioral Reviews*, *32*, 292–310. doi: 10.1016/j.neubiorev.2007.07.005

Speekenbrink, M., Lagnado, D. A., Wilkinson, L., Jahanshahi, M., & Shanks, D. R. (2010). Models of probabilistic category learning in Parkinson's disease: Strategy use and the effects of L-dopa. *Journal of Mathematical Psychology*, *54*, 123–136. doi: 10.1016/j.jmp.2009.07.004

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, *139*, 266–298.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*, 232–243.

Swainson, R., SenGupta, D., Shetty, T., Watkins, L. H. A., Summers, B. A., Sahakian, B. J., . . . , Robbins, T. W. (2006). Impaired dimensional selection but intact use of reward feedback during visual discrimination learning in Parkinson's disease. *Neuropsychologia*, *44*, 1290–1304. doi:10.1016/j.neuropsychologia.2006.01.028

Tharp, I. J., & Pickering, A. D. (2009). A note on DeCaro. Thomas, and Beilock (2008): Further data demonstrate complexities in the assessment of information-integration category learning. *Cognition*, *111*, 410–414. doi: 10.1016/j.cognition.2008.10.003

Trigg, J., & Kalish, M. (2011). Explaining how the mind works: On the relation between cognitive science and philosophy. *Topics in Cognitive Science*, *3*, 399–424.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category systems. *Psychonomic Bulletin & Review*, *8*, 168–176.

Wilkinson, L., Lagnado, D. A., Quallo, M., & Jahanshahi, M. (2008). The effect of feedback on non-motor probabilistic classification learning in Parkinson's disease. *Neuropsychologia*, *46*, 2683–2695. doi: 10.1016/j.neuropsychologia.2008.05.008

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