Title
Working Memory Capacity and Generalization in Predictive Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Wills, Andy
Barrasin, Thomas
McLaren, Ian

Publication Date
2011

Peer reviewed

Working Memory Capacity and Generalization in Predictive Learning

Andy J. Wills (a.j.wills@ex.ac.uk)
Psychology, University of Exeter,
Perry Road, Exeter EX4 4QG. UK

Thomas J. Barrasin (tombarrasin@hotmail.com)
Psychology, University of Exeter,
Perry Road, Exeter EX4 4QG. UK

Ian P. L. McLaren (I.P.L.McLaren@ex.ac.uk)
Psychology, University of Exeter,
Perry Road, Exeter EX4 4QG. UK

Abstract

The relationship between working memory and deliberative processing was examined in a human contingency learning experiment that employed the combined positive and negative patterning procedure of Shanks and Darby (1998). Participants with a large working memory capacity showed generalization consistent with the application of an opposites rule (i.e., a compound and its elements signal opposite outcomes), whilst individuals with a small working memory capacity showed generalization consistent with surface similarity. Working memory capacity was assessed via the Operation Span task (Turner & Engle, 1989). Implications for associative, inferential, and dual-process accounts of human learning are discussed.

Keywords: rules; associative learning; generalization; working memory; deliberative processing.

Introduction

The distinction between deliberative and non-deliberative processing, under a variety of different names, is fundamental to the study of cognition. For example, theorists seek to distinguish between propositional and associative learning (Mitchell, De Houwer & Lovibond, 2009), between analytic and nonanalytic categorization (Brooks, 1978), between automatic and intentional retrieval from memory (Jacoby, 1991), and between intuitive and deliberate reasoning (Kahneman, 2003). Deliberative processing is generally considered to be characteristic of thought processes that go beyond surface similarity to extract casual (De Houwer & Beckers, 2003) or abstract (Shanks & Darby, 1998) structure, thought processes that go beyond simple familiarity to episodic recollection (Jacoby, 1991), thought processes that are able to detect and correct irrational non-deliberative inferences (Kahneman, 2003). Deliberative thought processes are also often considered to be those that involve a degree of recurrence – in the sense that one goes through a series of intermediate stages to arrive at the final response (Milton & Wills, 2004). Another, related, way of capturing this idea of recurrence is to say that deliberative thought approximates the operation of a physical symbol system (Newell, 1980) – the ideas are related because certain recurrent, neural-like, structures have been shown to be able to implement a Universal Turing Machine (Siegelmann & Sontag, 1995).

In the current study we investigated the relationship between the availability of working memory resources and the extent to which people engage in deliberative processing when acquiring new information. People with comparatively large working memories learn some tasks more quickly (e.g. learning to trace electrical signals through logic gates; Kyllonen & Stephens, 1990), and other tasks more slowly (e.g. acquisition of a hard-to-verbalize category structure; De Caro, Thomas & Beilock, 2008; but see Tharp & Pickering, 2009), than people with comparatively small working memories. In the current study, we were interested primarily in the relationship between availability of working memory resources and the nature of what was learned.

One related study is that by De Houwer and Beckers (2003). In their forward cue competition experiment, participants first observed (in a computer game scenario) that firing a particular weapon (A) was followed by the destruction of a tank. Later on, weapon A was fired simultaneously with a new weapon, B. This compound firing also led to the destruction of the tank. They were then asked about the causal status of weapon B with respect to the destruction of the tank. On one, non-deliberative, account weapon B and the destruction of the tank have been repeatedly paired and thus one might say weapon B causes destruction of the tank by the mere fact of contiguity. However on another, deliberative, account one might argue that the causal status of B is uncertain because A causes destruction of the tank on its own, and B has never been used on its own. De Houwer and Beckers found that the imposition of a concurrent working memory load led to higher ratings for the extent to which B was considered to cause the tank's destruction, compared to a situation where the same contingencies were observed in the absence of a concurrent load. Their conclusion was that the imposition of a concurrent load interfered with the deliberative (deductive reasoning) processes required to work out that the causal status of B was uncertain, despite the fact it had been repeatedly paired with the destruction of the tank.

A.

Training			Test		
A+	B+	AB-	A?	B?	AB?
C-	D-	CD+	C?	D?	CD?
E+	F+	EF-	E?	F?	EF?
G-	H-	GH+	G?	H?	GH?
I+	J+		I?	J?	**IJ?**
		KL-	**K?**	**L?**	KL?
M-	N-		M?	N?	**MN?**
		OP+	**O?**	**P?**	OP?

B.

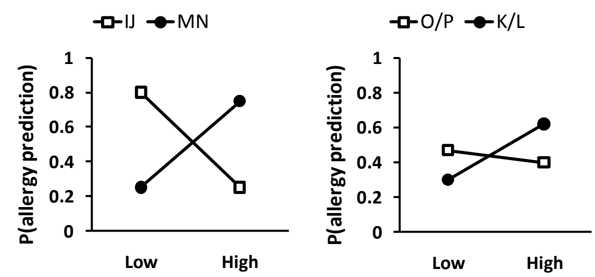

Figure 1: A. The training and test trial types in the Shanks and Darby (1998, Experiment 2) allergy prediction task; letters indicate foods eaten by a hypothetical patient Mr. X, + = patient develops an allergic reaction; - = patient does not develop an allergy reaction; ? = no feedback given. B. Critical test trials of Shanks and Darby (1998, Experiment 2) – probability of participants predicting an allergic reaction in Mr. X to novel meals, as a function of accuracy in the training phase.

In another related study, Waldron and Ashby (2001) demonstrated that concurrent working memory load retarded the acquisition of category structures definable in terms of a simple (single-attribute) rule, whilst the acquisition of category structures for which the rule was complex and non-intuitive was not significantly affected by concurrent load. Waldron and Ashby concluded that concurrent load interfered with the deliberative (rule-based) categorization process that would normally dominate in the simple-rule case, but that concurrent load left unaffected the non-deliberative processes underlying the acquisition of the more complex category structure.

One of the inherent difficulties in attempting to demonstrate a relationship between the availability of working memory resources and the extent to which learning proceeds deliberatively is to find a type of learning behavior that is unambiguously outside the scope of non-deliberative theories of cognition. For example, Nosofsky and Kruschke (2002) have argued that the results of Waldron and Ashby (2001) can be accounted for by a non-recurrent, non-deliberative, exemplar model (ALCOVE; Kruschke, 1992). The basis of Nosofsky and Kruschke's argument is that concurrent load may be hypothesized to disrupt learned selective attention. This disruption will affect learning tasks in which selective attention is helpful (such as categories defined by a single attribute). Forward cue competition is also, at least in part, the result of learned selective attention (Kruschke, Kappeman & Hetrick, 2005; Wills, Lavric, Croft & Hodgson, 2007), so the idea that concurrent load disrupts selective attention might also, in principle, account for the forward cue competition results of De Houwer and Beckers (discussed above).

In the current studies, we examined the role of working memory in the learning and generalization task introduced by Shanks and Darby (1998). The task is unusual in that the performance of a subset of participants in this task is widely considered to be strong evidence for the role of deliberative processing in learning. This view is held both by those who argue for a central role of non-deliberative (associative) processes in human learning (e.g. Cobos, Almaraz & Garcia-Madruga, 2003), and for those who argue against this position (e.g. Mitchell et al., 2009). Verguts and Fias (2009) have argued that the behavior of this subset of participants can be accounted for by a recurrent connectionist model – this argument is consistent with the position, outlined above, that recurrence is a characteristic feature of deliberative processing and that recurrent network architectures can implement general-purpose computational systems.

The design and key result of Shanks and Darby (1998) is shown in Figure 1. Participants were asked to take the role of an allergist, attempting to predict which foods will cause an allergic reaction in a hypothetical patient, Mr. X. In Figure 1, letters stand for foods, "+" indicates the presence of an allergic reaction, and "-" indicates the absence of an allergic reaction. The training phase contained two complete positive patterning problems (e.g. A+, B+, AB-) and two complete negative patterning problems (e.g. C-, D-, CD+). Training also contained four incomplete patterning problems (e.g. participants see I+ and J+ but not IJ). The critical results concern participants' generalization to novel items, such as IJ, in the absence of feedback. For example, say you have observed that Mr. X develops an allergic reaction when he eats ice cream (I) and when he eats jelly (J). Do you predict the presence or absence of an allergic reaction when eating ice cream and jelly together?

A non-deliberative, surface similarity, process is likely to predict allergic reaction to IJ, as IJ is similar to both I and J, both of which produced an allergic reaction. A deliberative process, however, might detect that an opposites rule succinctly captures the information available during training – single foods produce the opposite reaction to their compounds. On this basis, IJ is predicted to not result in an allergic reaction, because this is the opposite outcome to that for I occurring on its own (and for J occurring on its own). As shown in Figure 1, Shanks and Darby found that participants who achieved a high level of accuracy during training showed generalization consistent with the application of an opposites rule, while participants who performed less well in training showed generalization consistent with surface similarity. Our hypothesis is that this transition in generalization reflects a transition from non-deliberative to deliberative processing during the course of training. We further hypothesize that, if opposites-rule and surface-feature generalization are indeed the products of deliberative and non-deliberative processing respectively, then the availability of working memory resources should determine whether opposites-rule or surface-feature generalization is seen (under the assumption that deliberative processing makes greater demands on working memory than non-deliberative processing).

Existing evidence could be employed, in a fairly indirect manner, to argue either for, or against, our hypothesis. On the one hand, Winman, Wennerholm, Juslin and Shanks (2005) demonstrated that opposites-rule generalization was related to performance on Raven's Progressive Matrices (RPM). RPM are considered to be a measure of general intelligence (*g*) and *g* appears to be related to working memory capacity (Conway, Kane & Engle, 2003). Hence one might argue that those with high working memory capacity should be more likely to show opposites-rule generalization than those with low working memory capacity. On the other hand, De Houwer and Vandorpe (2009) demonstrated performance consistent with opposites-rule generalization in the Implicit Association Test (IAT; Greenwald, McGhee & Schwartz, 1998). Although a matter of some debate (Fazio & Olson, 2003), the IAT (as its name suggests) is often considered to index non-deliberative processing. Also, opposites-rule generalization is related to participants demonstrating an inverse base-rate effect (Winman et al., 2005), yet demonstration of an inverse base-rate effect has been reported to be unaffected by a concurrent load (Lamberts & Kent, 2007).

In the current article, we examined the relationship between working memory and deliberative processing in two ways. We measured individuals' working memory using an Operation Span task (OSPAN; Turner & Engle, 1989) and tested the hypothesis that those with relatively large working memory capacity would show generalization more consistent with the application of an opposites rule, whilst individuals with a relatively small working memory capacity would show generalization more consistent with surface similarity.

Experiment

Method

Participants and apparatus. Forty-two adults from the Exeter and Guernsey regions of the United Kingdom took part on a voluntary basis. They were tested individually in a quiet testing room using a PC laptop (17" screen) running the E-prime software package (Version 1.1, Psychology Software Tools, Pittsburgh, USA).

Design and stimuli (learning task). The design of the learning task is shown in Figure 1A. For half the participants the foods A-P were, respectively, coconut, cheese, apple, orange, carrots, cabbage, chips, nuts, eggs, banana, beetroot, rice, milk, and garlic. For the remaining participants, the foods assigned to A and B were swapped with those assigned to C and D, and similarly for E/F and G/H, for I/J and K/L, and for M/N and O/P.

Procedure. Participants were asked to assume the role of an allergist, predicting whether a hypothetical patient, Mr. X, would or would not develop an allergic reaction after eating a meal containing certain foods. On each trial, food names were presented on the screen, and participants pressed a key to indicate whether or not Mr. X would suffer an allergic reaction. No time limit was set for these responses. During the training phase, each trial was followed by a feedback message of 1500ms duration (e.g. "Correct! Mr. X developed an allergic reaction"). No feedback was given during the test phase. The training phase comprised eight blocks; each block contained two of each of the 18 training trial types shown in Figure 1A, presented in a random order. The test phase comprised two of each of the 24 test trial types shown in Figure 1A, again presented in a random order. The transition between blocks and phases was not signaled to participants, although they were forewarned that feedback would not be available towards the end of the experiment.

After a short break (1-2 minutes), the participants proceeded to the operation span (OSPAN) task. In this task, participants were presented with a total of 60 trials split into 15 groups of trials. There were 3 groups each of 2, 3, 4, 5 and 6 trials and participants were presented with the groups of trials in a pseudo-random order. Each trial consisted of a simple mathematical equation (e.g. $2 \times 3 + 1 = 7$) presented simultaneously with a word (e.g. BED). The participant's task was to indicate whether the answer to the equation was correct or incorrect. They were also required to remember each word. Each equation/word combination was response terminated with a timeout of 5 seconds. At the end of each group of trials, participants were asked to recall, in order, the words presented within that group. For example, in a block size of 3, participants would be presented with three equation/word pairs. After the presentation of the third pair, participants would be asked to recall the first word in the group followed by the second word and finally, the third word. Participants were not allowed to backtrack and change a previously given answer. There was no limit placed on time for recall, and no feedback was given. A participant's score on the OSPAN task was calculated as the sum of correctly recalled words for trial groups that were perfectly recalled (e.g. Conway & Engle, 1994; Turner & Engle, 1989).

The operation span task proper was preceded by three practice spans (all of length two) to familiarize participants with the task. These practice spans were not analyzed.

Results

Our hypothesis was that those with a high working memory capacity (as measured by the OPSAN task) would show generalization consistent with an opposites rule, whilst participants with a low working memory capacity would show generalization consistent with surface similarity. High and low working memory capacity was operationalized as the upper and lower quartiles of the sample on the OSPAN score (Conway & Engle, 1994).

Figure 2 (left panel) illustrates the generalization to novel compounds MN and IJ, as a function of high vs. low OSPAN. An ANOVA with one between-subjects factor (high vs. low span), and one within-subjects factor (stimulus; MN vs. IJ), revealed a significant interaction between these factors, $F(1,20) = 15.87$, $p = 0.001$. The main effects of span and stimulus were not significant, $F(1,20) < 2.55$, $p > 0.10$. Under an opposites rule, the appropriate response to MN is that an allergic reaction is expected, and the appropriate response to IJ is that an allergic reaction is not expected. Under the application of surface similarity, the predictions are reversed. The generalization to novel compounds MN and IJ shown by high working memory capacity participants is therefore more consistent with the application of an opposites rule, whilst the generalization shown by low working memory capacity participants is more consistent with the application of surface similarity. Figure 2 (right panel) illustrates a similar result for novel test stimuli K/L and O/P although, as in Shanks and Darby (1998), the interaction is marginally significant, $F(1,20) = 3.91$, $p = 0.06$. The main effects are non-significant, $F(1,20) < 1$.

Performance on familiar test items (i.e. those also seen during training) was consistent with the feedback received during training, for both groups. High OSPAN participants predicted an allergic reaction on 92% of occasions for stimuli that had been associated with allergy in training, and predicted an allergic reaction on 13% of occasions for stimuli that had not been associated with allergy in training. For low OSPAN participants, the figures were 71% and 39%. An ANOVA with one within-subjects factor (stimulus type: allergy vs. no allergy) and one between-subjects factor (high vs. low working memory) revealed a main effect of stimulus type, $F(1, 20) = 67.3$, $p < 0.0005$, and a significant interaction, $F(1, 20) = 11.9$, $p = 0.003$. There was no main effect of group, $F(1,20) < 1$. The significant interaction reflects more accurate performance on familiar items by the high working memory capacity participants.

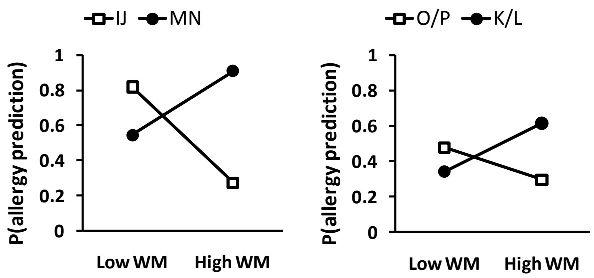

Figure 2: mean probability of participants predicting an allergic reaction in response to novel stimuli IJ, MN, O/P (the mean of responses to O and P), and K/L, shown as a function of participants' working memory capacity (upper vs. lower quartile).

Discussion

We found that participants with a comparatively large working memory capacity generalized to novel stimuli in a learning task (Shanks & Darby, 1998) in a way consistent with the application of an abstract rule, whilst participants with a comparatively small working memory capacity generalized to novel stimuli in a way consistent with surface similarity. Those participants with a large working memory capacity also learned more quickly than those with a small working memory capacity. Both positive (Kyllonen & Stephens, 1990), and negative (DeCaro et al., 2008) relationships between working memory capacity and rate of learning have been reported in other tasks.

The learning task employed in these studies was introduced by Shanks and Darby (1998). The task is unusual in that there is a broad consensus, from a range of theoretical perspectives, that the abstract-rule generalization seen in this task is outside the scope of non-deliberative thought processes (e.g. those processes captured by simple associative, and non-recurrent connectionist, models; Cobos et al., 2003; De Houwer & Beckers, 2003; Mitchell et al., 2009; Verguts & Fias, 2009). The consensus concerning the Shanks and Darby task stands in contrast to the contentious nature of some other forms of putatively deliberative behavior that have been reported in adults (e.g. DeCaro et al., 2008 vs. Tharp & Pickering, 2009).

The explanation we tentatively offer for our results is that participants initially approach the training phase through a process of exemplar storage and retrieval. This is a relatively non-deliberative process (which is not to say it is necessarily entirely automatic). Later in training, some participants notice there is a non-intuitive rule that substantially reduces the number of things one has to remember. At limit, all one needs to remember is the rule that compounds predict the opposite to their elements, and that the compounds CD, EF, and KL make Mr. X sick. Everything else can be derived – for example, A on its own will make Mr.X sick, because A is not in any of the three compounds that make him sick (CD, EF, KL), and compounds predict the opposite to their elements.

This process of rule extraction is assumed to be relatively deliberative and effortful, requiring as it does the generation of verbal hypotheses and then testing them against subsequently presented training items. If the process of rule extraction is assumed to be relatively deliberative and effortful, then it is not unreasonable to assume it might be more likely to occur in those with a relatively large working memory. Generalization to novel test items will depend on the nature of the representations developed during training. For those with a small working memory capacity, the representations are exemplars, and generalization would therefore be expected to be on the basis of surface similarity. Those with a large working memory, however, extract the opposites rule, and generalization might therefore be expected to be on the basis of that opposites rule.

The idea that rule extraction in the Shanks-Darby task is deliberative, effortful, and requiring of working memory resources, is also supported by related work in our laboratory. Across two experiments, Wills, Graham, Koh, McLaren and Rolland (2011) demonstrated that the imposition of a concurrent working memory load during the training phase of the Shanks-Darby task resulted in similarity-based generalization during the test phase. In the absence of concurrent working memory load during training, participants produced opposites-rule generalization. Opposites-rule generalization in the absence of load was to be expected in these experiments as participants were trained to a high criterion (89% accuracy); those not meeting the criterion were excluded from analysis. Interestingly, Wills et al. (2011) found that the presence of concurrent working memory load during the test phase had no effect. Hence, it appears to be the extraction of a rule in this task that requires substantial working memory resources, rather than the application of an extracted rule. It may be that substantial working memory resources are required to generate and/or evaluate a rule that summarizes (and hence simplifies) the information presented during training.

The explanation we offer for our results falls amongst the broad class of explanations that assume cognition in adult humans is the product of at least two systems – one system that is deliberative, and perhaps approximates the function of a physical symbol system (Newell, 1980), and another system that is non-deliberative, and which might be approximated by a simple associative system. Explanations of this general class include those forwarded by Ashby et al. (1998), Brooks (1978), and Sloman (1996).

In addition to dual-process accounts of human cognition, another class of theory is that all learning is the product of a deliberative system (e.g. inferential accounts; Mitchell et al., 2009). In relation to the results of the current experiments, inferential accounts would presumably assume that not only opposites-rule performance, but also surface similarity performance, was the product of an inferential process in this task. In order to predict any effect of working memory capacity, such an account must assume that opposites-rule performance is more effortful than surface similarity performance – perhaps because participants arrive with a pre-experimental hypothesis that similar meals lead to similar outcomes, whilst the non-intuitive opposites-rule hypothesis is only arrived at through a relatively effortful process of hypothesis-testing during training. Having a relatively limited working memory capacity presumably interferes with this process, leaving the participant with just memory for the examples (which is employed for familiar test items) and a pre-experimental surface similarity hypothesis (which is employed for novel test items).

In summary, an inferential explanation seems to need to assume both the presence of a relatively non-deliberative exemplar storage and retrieval process, and that certain types of inferential process are also relatively non-deliberative (e.g. inferences on the basis of a pre-experimental hypothesis about novel test items). In other words, an inferential explanation seems to need to assume the presence of not only relatively non-effortful forms of learning and retrieval, but also the presence of relatively non-effortful forms of inference. When expressed in those terms, an inferential account seems to largely converge with the dual-process we offer above.

In summary, we have reported an experiment that suggests the availability of working memory resources is an important determinant of how we generalize from what we have learned. In our studies, information learned and tested in the presence of substantial working memory resources seems to lead to generalization more consistent with the application of an abstract rule, whilst information learned and tested in the presence of more limited working memory capacity seems to lead to generalization more consistent with surface similarity.

Acknowledgments

This research was supported by British Council grant RC105. The authors thank Jan De Houwer, David Shanks, and Tom Verguts for their helpful comments, and Chris Longmore and Charlotte Reilly for their assistance. This article is dedicated to Lee Brooks 1938 – 2010.

References

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*, 442-481.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization.* Hillsdale, N.J.: Erlbaum.

Cobos, P. L., Almaraz, J., & Garcia-Madruga, J. A. (2003). An associative framework for probability judgment: An application to biases. *Journal of Experimental Psychology-Learning Memory and Cognition, 29*(1), 80-96.

Conway, A. R. A., & Engle, R. W. (1994). Working-Memory and Retrieval - a Resource-Dependent Inhibition Model. *Journal of Experimental Psychology-General, 123*(4), 354-373.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547-552.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Somtimes less working memory capacity is better than more. *Cognition, 107*(1), 284-294.

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *Quarterly Journal of Experimental Psychology, 56B*(4), 345-357.

De Houwer, J., & Vandorpe, S. (2009). Using the Implicit Association Test as a Measure of Causal Learning Does not Eliminate Effects of Rule Learning. *Experimental Psychology, 57*(1), 61-67.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Jacoby, L. L. (1991). A Process Dissociation Framework - Separating Automatic from Intentional Uses of Memory. *Journal of Memory and Language, 30*(5), 513-541.

Kahneman, D. (2003). A perspective on judgment and choice - Mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

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

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology-Learning Memory and Cognition, 31*(5), 830-845.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2*(2), 129-160.

Lamberts, K., & Kent, C. (2007). No evidence for rule-based processing in the inverse base-rate effect. *Memory & Cognition, 35*(8), 2097-2105.

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

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*(2), 183-+.

Newell, A. (1980). Physical Symbol Systems. *Cognitive Science, 4*(2), 135-183.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review, 9*(1), 169-174.

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 24*(4), 405-415.

Siegelmann, H. T., & Sontag, E. D. (1995). On the Computational Power of Neural Nets. *Journal of Computer and System Sciences, 50*(1), 132-150.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin, 119*, 3-22.

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*(3), 410-414.

Turner, M. L., & Engle, R. W. (1989). Is Working Memory Capacity Task Dependent. *Journal of Memory and Language, 28*(2), 127-154.

Verguts, T., & Fias, W. (2009). Similarity and Rules United: Similarity- and Rule-Based Processing in a Single Neural Network. *Cognitive Science, 33*, 243-259.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8*(1), 168-176.

Wills, A.J., Graham, S., Koh, Z., McLaren, I.P.L. and Rolland, M.D. (2011). Effects of Concurrent Load on Feature- and Rule-based Generalization in Human Contingency Learning. *Journal of Experimental Psychology: Animal Behavior Processes.* Advance online publication.

Wills, A. J., Lavric, A., Croft, G. S., & Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: Evidence from event-related potentials and eye tracking. *Journal of Cognitive Neuroscience, 19*(5), 843-854.

Winman, A., Wennerholm, P., Juslin, P., & Shanks, D. R. (2005). Evidence for rule-based processes in the inverse base-rate effect. *Quarterly Journal of Experimental Psychology, 58A*(5), 789-815.