Title
Connecting rule-abstraction and model-based choice across disparate learning tasks

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Don, Hilary J
Goldwater, Micah B
Otto, A Ross
et al.

Publication Date
2015

Peer reviewed

Connecting rule-abstraction and model-based choice across disparate learning tasks

Hilary J. Don (hdon7006@uni.sydney.edu.au)[1]
Micah B. Goldwater (micah.goldwater@sydney.edu.au)[1]
A. Ross Otto (rotto@nyu.edu)[2]
Evan J. Livesey (evan.livesey@sydney.edu.au)[1]
[1]School of Psychology, Brennan McCallum Building, University of Sydney, NSW 2006 AUS
[2]Center for Neural Science, New York University, New York, NY 10003 USA

Abstract

Recent research has identified key differences in the way individuals make decisions in predictive learning tasks, including the use of feature- and rule-based strategies in causal learning and model-based versus model-free choices in reinforcement learning. These results suggest that people rely to varying degrees on separable psychological processes. However, the relationship between these types of learning strategies has not been explored in any depth. This study investigated the relationship between feature- vs rule-based strategies in a causal learning task and indices of model-free and model-based choice in a two-step reinforcement learning procedure. We found that rule-based transfer was associated with the use of model-based, but not model-free responding in a two-step task.

Keywords: predictive learning; individual differences; rule vs. feature generalization; model-based vs. model-free; cognitive control; associative learning; decision-making

Introduction

Theories of learning and decision making often assume a contribution from multiple distinct processes (Mitchell, De Houwer & Lovibond, 2009; Balleine & O'Doherty, 2010; Jacoby, 1991; Kahneman, 2011). Although these processes have been defined in a range of different ways, they tend to include one process that requires cognitive control and deliberate thought and one that is simpler and relatively automatic. The former process tends to be described as effortful and rule-based, extracting causal or abstract structure from the environment in order to plan behavior (De Houwer & Beckers, 2003). The latter process is considered by many to be based on associative mechanisms, with responding to novel stimuli operating on the basis of surface similarity or featural overlap (McLaren et al., 2014).

Typically, research in these areas entails presenting a series of trials in which participants learn to predict relationships between cues and outcomes, or actions and outcomes. It is often difficult to distinguish between the contributions of distinct processes, as they result in very similar behavior in most circumstances, and are sometimes examined under conditions that favor a particular process (e.g. Waldron & Ashby, 2001). Nevertheless, in recent years, a range of tasks that use carefully designed analyses of training and transfer items have been successful in identifying separable response strategies that suggest the involvement of distinct psychological processes in both measuring generalization across stimuli and reward-driven choice (e.g., Shanks & Darby, 1998; Daw, Niv & Dayan, 2005, respectively). These tasks originate from different but conceptually similar lines of research, in which there is some evidence that behavior takes into consideration abstract structure in the planning of goals, as well as evidence of behavior that is consistent with the formation of simple associations. Despite the clear similarity in the distinctions that are drawn using these tasks, and the apparent presence of individual differences across participants, the relationship between these tasks has received very little attention. We will describe two such tasks that are relevant to the current study.

The first concerns the generalization of learned information to novel stimuli (Shanks & Darby, 1998). Participants were asked to assume the role of a doctor whose task was to determine which foods were causing an allergic reaction in their fictitious patient, Mr X. Within this scenario, participants learned about several food-reaction (cue-outcome) relationships in a sequential trial-and-error fashion, before being presented with the critical test phase. The design of Shanks and Darby (1998) is shown in Table 1.

Table 1: Patterning task design

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?

Note: Letters A-P represent randomly allocated foods used as cues. These cues were followed by an allergic reaction (+) or no allergic reaction (-). Critical transfer trials are depicted in bold.

Participants were trained with two complete negative patterning discriminations, in which two food cues (e.g. A and B) each cause an allergic reaction outcome (+) when they are eaten individually but when eaten together, do not cause an allergic reaction (i.e. A+/B+/AB-). Participants were also trained with two complete positive patterning discriminations in which two cues that do not cause the outcome individually do result in the outcome when presented together (e.g. C-/D-/CD+). In addition, participants were presented with a number of cues that appeared either individually (e.g. I+/J+) or in compound (e.g. KL-) but not both.

Accurate performance on these discriminations can be achieved through learning the associations between combinations of cues and outcomes. However the structure of the task can also be described by an abstract "opposites" rule. That is, individual cues and their compounds predict opposite outcomes. In the test phase, participants continued to predict whether or not food cues would cause an allergic reaction, in the absence of feedback. This phase included all the training cues, as well as the remaining cues from the incomplete discriminations (e.g. IJ). Participants' responses to these novel transfer cues were of primary interest. Generalisation based on surface similarity would predict an "allergic reaction" response to IJ, due to its similarity to I and J. On the other hand, generalisation based on extraction and application of the opposites rule would predict a "no reaction" response to IJ. This pattern of feature- and rule-based generalisation is illustrated in Figure 1.

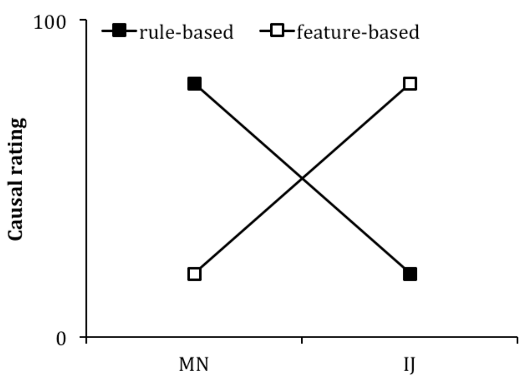

Figure 1. Predicted outcome ratings for MN and IJ transfer trials for rule-based and feature-based generalization.

The second task aims to dissociate model-free and model-based strategies in reinforcement learning, which each determine how actions are evaluated from previous experiences (Daw et al., 2005). A model-free strategy repeats actions that have previously been rewarded, consistent with associative principles. A model-based strategy takes into account a model of the environmental structure, reasoning about action values and current goals in order to plan behavior. Recently, Daw et al. (2011) used a sequentially structured choice task in order to dissociate these processes. In the two-step task, a first-stage binary choice (A1 vs. A2) led probabilistically to a second-stage state (S1 vs. S2), in which a second choice (A3 vs. A4; A5 vs. A6) resulted in either reward, or no reward. Each of the first-stage choices led to a particular second-stage state (e.g. A1-S1; A2-S2) 70% of the time (common transitions), and to the other second-stage state (e.g. A1-S2; A2-S1) 30% of the time (rare transition).

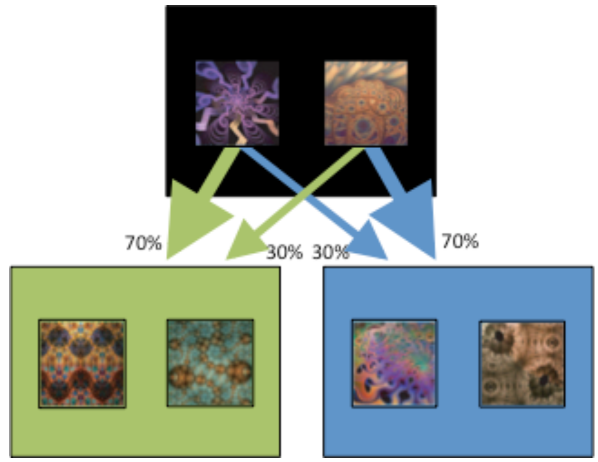

Figure 2. Two-step task transition structure. Each first stage choice leads to one of the second-stage states 70% of the time. The probability of receiving reward on each of the second-stage states changed slowly over the course of the experiment.

To ensure participants continually searched for the optimal action, the probability of receiving a reward on each of the second-stage choices changed slowly over the course of the experiment. The critical dependent measure is the likelihood of participants repeating the same first-stage choice on each trial based on the previous trial's outcomes. Take, for example, a choice that results in a rare transition to a second-stage state (e.g. A1-S2), in which a rewarded choice is made. A model-free strategy predicts that the participant should repeat that first-stage choice action, as it ultimately resulted in reward (Figure 3A). Conversely, a model-based strategy predicts that the likelihood of repeating the same choice will decrease, as the value of the alternative choice that *commonly* leads to the rewarded second-stage state (C2) should increase. Model-based choice therefore requires participants to have learned both the second-stage reward probabilities and the transition structure of the task, and to use this information to prospectively plan subsequent first-stage choice. Thus, the hallmark of model-based responding is an interaction between reward and transition type on the previous trial on first-stage choice (Figure 3B). Daw et al. (2011) found a mixture of model-based and model-free behavioral contributions at a population level and within many individuals. However a number of participants showed responses consistent with purely model-free or purely model based behavior.

Across both task domains, the use of a particular strategy may be influenced by task conditions. For example, rule- and model-based processes that are more reliant on cognitive resources are reduced when participants are trained under a concurrent load (Wills et al., 2011, Otto, Gershman, Markman & Daw, 2013). Nevertheless, tacit in this research is the idea that individual differences in the degree to which participants employ each process may also be important. Individual differences in working memory capacity have been shown to predict performance on rule-based categorization tasks (DeCaro, Thomas & Beilock, 2008). Further, Shanks & Darby (1998) found that efficient learners during training were more likely to show rule-based generalization at test than inefficient learners, and their task has also been used to establish a relationship between rule-based transfer and rule-mediated processes in the inverse base-rate effect (Winman, Wennerholm, Juslin & Shanks, 2005).

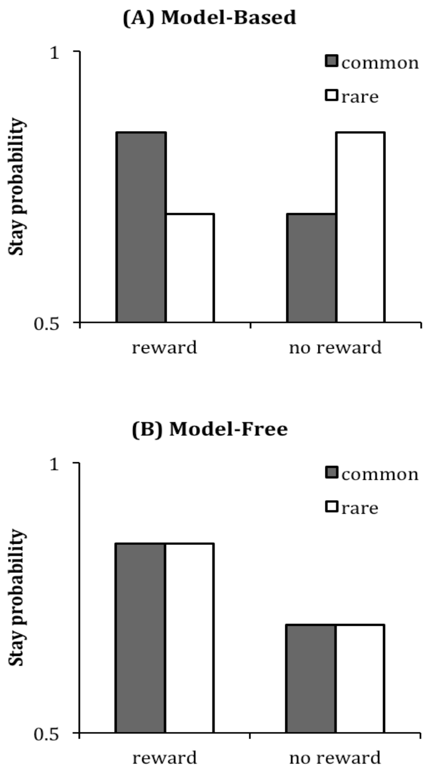

Figure 3. (A) A model-based choice strategy predicts that reward after rare transitions will influence the following first-stage choice, leading to an interaction between reward and transition type. (B) A model-free strategy predicts that a rewarded first-stage choice is more likely to be repeated regardless of whether reward occurred on a common or rare transition.

One advantage of both the patterning task and the two-step choice task is that separable processes predict qualitatively different patterns of results. Further, responses to critical items can neither be considered accurate nor inaccurate. Consequently, individual differences do not necessarily reflect better or worse performance, but rather a propensity to rely on a particular process, and so connections between them are not as simple as the degree to which subjects behaved non-randomly in both tasks (Otto, Skatova, Madlon-Kay, & Daw, 2015; but see Shanks & Darby, 1998). Likewise, previous research demonstrates a stable tendency within individuals to use rule-based vs. exemplar-based learning across multiple conceptual learning tasks, in the laboratory and the classroom (McDaniel, Cahill, Robbins & Wiener, 2014). Given the variety of dual-process theories within learning and cognition, determining whether there are relationships between tasks that purport to measure similar dissociable processes requires further consideration. With the exception of McDaniel et al. (2014), there has been little attempt to verify whether these tasks are measuring the same, or even related constructs. Furthermore, there may be important differences between the dissociations that these tasks reveal. We aim to take this form of research in a new direction, relating strategies for generalization to strategies for reward-driven choice.

Across theories of generalization and choice behavior it is appealing to conceptualize distinctions between reflective and associative processes as features of the same two general, independent systems. Previous research also suggests that individuals may be consistent in their tendency to engage a particular system across tasks (McDaniel et al., 2014). Two predictions that fall out of this connection are that, a) individuals who show rule-based generalization will also show model-based choice, and b) greater feature-based generalization may predict more model-free behavior. Our primary goal for the current experiment is to evaluate these two possibilities.

Method

Participants

Forty undergraduate psychology students from the University of Sydney participated in exchange for partial course credit (18 female, mean age = 19.98, *SD* = 4.05).

Apparatus and Stimuli

Experimental stimuli in the patterning task included 300 x 300 pixel images of *coffee*, *banana*, *fish*, *lemon*, *cheese*, *garlic*, *apple*, *eggs*, *peanuts*, *mushrooms*, *strawberry*, *milk*, *bread*, *avocado, broccoli, olive oil, cherries, butter, chocolate, carrots, peach, bacon, peas and prawns*. All images were presented on a white background, with accompanying labels in blue text. Foods were randomly allocated to cues A-P for each participant. In the two-step task, first- and second-stage choices were denoted by randomly allocated fractal images, presented on black (first-stage) or colored (second-stage) backgrounds. Participants were tested individually using a standard PC.

Procedure

Participants completed both the patterning task and two-step task in counterbalanced order.

Patterning Task In the patterning task, participants were asked to assume the role of a doctor whose task was to determine which foods were causing allergic reactions in their fictitious patient, Mr X. On each trial, participants were presented with one or two food cues on the upper half of the screen, and were required to predict whether an allergic reaction would occur by clicking either a “no allergic reaction” or “ALLERGIC REACTION” option beneath the food cues. Participants were instructed that at first they would have to guess, but that using the feedback provided, their accuracy should improve over time. When an outcome was selected, the options disappeared and feedback was provided while the food cues remained on the screen. The correct answer appeared, accompanied by either the word "CORRECT" in green, or "INCORRECT" in red, depending on the accuracy of the prediction. Each trial type (see Table 1) was presented twelve times during the training phase. The position of compound cues on screen was counterbalanced across the course of training (e.g. six presentation of AB, and six presentations of BA).

A test phase was administered immediately following training. Participants were instructed that in this phase they were required to use the knowledge they had gained in the previous phase. On each trial, one of the test items (see Table 1) was presented, and participants were asked to rate the likelihood that an allergic reaction would occur on a 10-point linear analogue scale ranging from "definitely WILL NOT occur" to "definitely WILL occur". Two blocks of the test phase were completed, with each test item presented once per block.

After completing the transfer phase, participants completed a manipulation check to assess explicit knowledge of relational rules. The first part was an open question asking participants to describe any general rule they may have noticed during the experiment. The second part required participants to answer two forced choice questions. As in Harris & Livesey (2008), participants in the patterning condition were asked:

Did you notice that if A predicted an allergic reaction, and B predicted an allergic reaction, then the combination of A and B predicted no allergic reaction? (negative patterning) and,

Did you notice that if A predicted no allergic reaction, and B predicted no allergic reaction, then the combination of A and B predicted an allergic reaction? (positive patterning).

Two-Step Task Participants completed 200 trials of the two-step choice task (Figure 2). On each trial, two fractal images representing the first-stage options appeared on a black background. Participants were required to make a left or right response using the "Z" or "?" key, respectively. Once a choice was made, the background changed to either blue or green to indicate the second-stage state, and the selected first-stage image moved to the top of the screen. Another two fractal images were presented and participants were again required to make a choice response. Feedback was then provided while the selected image remained highlighted on screen. Participants were presented with either an image of a coin (reward), or the number zero (no reward).

Results

Following previous research (Otto Raio et al., 2013; Otto, Gershman et al., 2013), one participant was excluded for missing greater than 15 response deadlines in the two-step task, and two participants were excluded for showing no reward sensitivity at the second-stage level, i.e. $P(\text{stay}|\text{win}) < 50\%$. Thirty-seven participants remained in the following analyses.

Patterning Task

Analysis of the patterning task focused on the compound transfer cues, as these provide the clearest and most interpretable test of the feature- and rule-based distinction. The difference in causal ratings for MN and IJ (MN-IJ) was interpreted as a measure of generalization. This resulted in a score ranging from -100 – 100. A high score indicated greater rule-based transfer (high rating for MN, low rating for IJ), while a low score indicated greater feature-based transfer (high rating for IJ, low rating for MN). Twenty participants had a negative transfer score, revealing a pattern of responses consistent with generalization based on surface similarity. Seventeen participants had positive transfer scores, suggestive of generalization on the basis of the abstract patterning rule. The distribution of scores is shown in Figure 5. In the manipulation check, 28 (out of 37) participants verbalized either a general opposites rule, or both the positive and negative patterning rules. A further four participants verbalized only the negative patterning rule, and two participants verbalized only the positive patterning rule. Thirty-six participants were able to identify one or both of the patterning rules in the forced choice questions. The use of rule transfer was not significantly correlated with the ability to verbalize ($r = .294$, $p = .077$), or identify ($r = .049$, $p = .774$), a patterning rule.

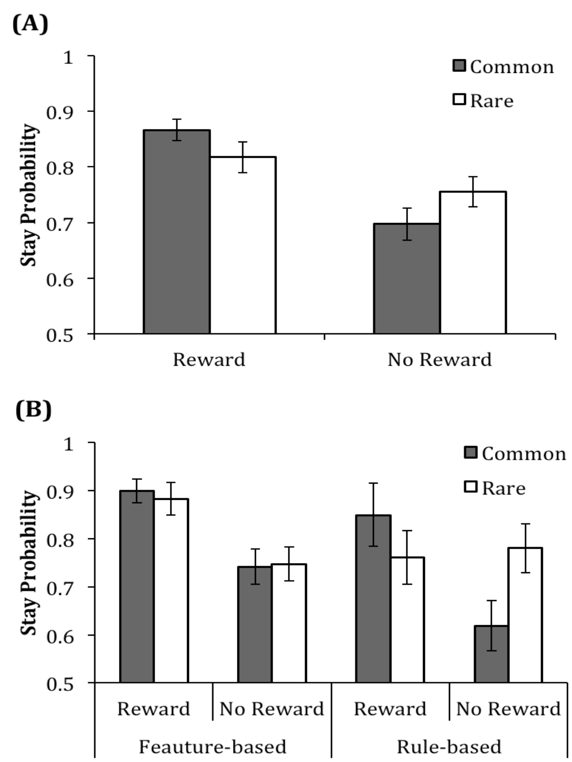

Figure 4. Probability of repeating a first-stage response in the two-step task for (A) all participants and (B) participants using feature- and rule-based generalization in the patterning task.

Two-Step Task

Figure 4A shows the effects of reward and transition type on first-stage outcome choice for all participants in the two-step task. We estimated a mixed effects logistic regression (Pinheiro & Bates, 2000) with first-stage choice (stay vs. switch) as the dependent variable, using binary predictors that indicated whether a reward was received on the previous trial, and transition type on the previous trial (common vs. rare). Full coefficient estimates are reported in Table 2. There was a significant main effect of reward, revealing a tendency to repeat rewarded first-stage choices ($p < .001$). The interaction between reward and transition type suggests a significant model-based contribution to choice ($p < .001$).

Table 2. Logistic regression coefficients indicating the influence of previous trial outcome, previous trial transition type, and patterning transfer on first-stage choice repetition.

Predictor	Estimate (SE)	*P* value
Intercept	1.63 (.15)	< .001*
Reward	0.55 (.08)	< .001*
Transition Type	0.07 (.04)	.093
Transfer	-0.002 (.002)	.316
Reward x Transition Type	0.23 (.06)	< .001*
Reward x Transfer	-0.001 (.001)	.258
Transition Type x Transfer	-0.0003 (.0005)	.618
Reward x Transition Type x Transfer	0.002 (.0007)	.024*

Relationship Between Tasks

To illustrate the relationship between patterning transfer and performance on the two-step task, we plotted the relationship between raw transfer scores and an index of model-free and model-based responding for each participant (Figure 5). This index was computed by taking individual participants' coefficients for reward, and reward x transition type interaction, respectively. Statistically, including z-scored transfer scores as a predictor in the logistic regression revealed no significant interaction between transfer and reward, suggesting that there was no relationship between patterning transfer and model-free responding ($p = .258$). However, there was a significant three-way interaction between reward, transition type and transfer, which indicates that higher transfer scores were associated with greater model-based responding ($p = .024$). To further illustrate this interaction, Figure 4B shows the probability of repeating a first-stage response for participants using feature- and rule-based transfer in the patterning task. The lowest third of transfer scores were considered highly feature-based, ($n = 12$; $M = -87.01$), while the top third of transfer scores were considered highly rule-based ($n = 12$; $M = 85.33$).

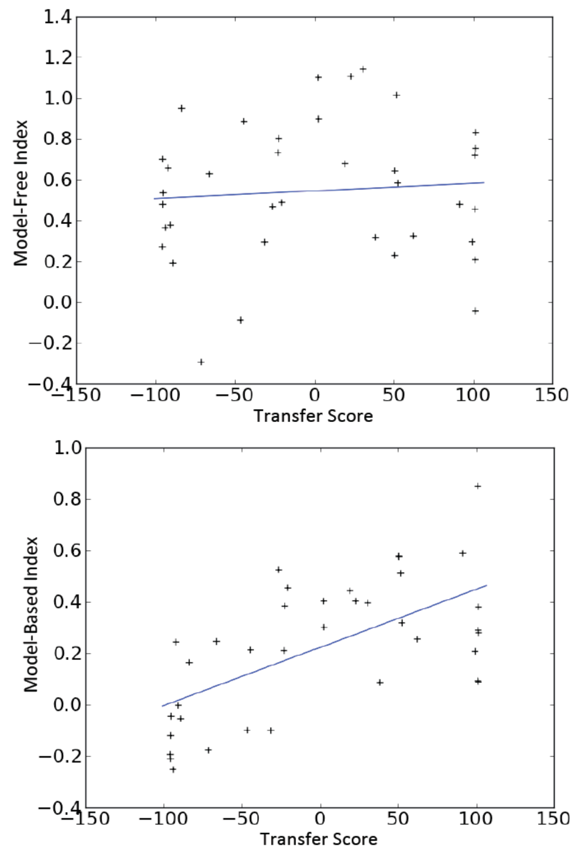

Figure 5. Scatterplots showing the relationship between the patterning rule transfer score on the x-axis and an index of model-free (top panel) and model-based (bottom panel) choice, estimated from the logistic regression, in the two-step task on the y-axis.

Discussion

This study examined the relationship between separable processes in two predictive learning paradigms. Individual differences in patterns of responding were identified in a patterning task, and a two-step sequential choice task, which may reflect the use of either effortful, rule-based processes, or simple associative or feature-based processes. In the two-step task, we observed evidence of both model-free and model-based behavior on a group-level, which is consistent with previous findings (Daw et al., 2011; Otto, Raio et al.; 2013; Otto, Gershman et al., 2013).

Importantly, performance on the patterning task was significantly related to choice behavior on the two-step task. Namely, generalization had a predictive relationship specific to model-based, but not model-free choice. Participants who were able to extract and apply the abstract patterning rule to novel compounds exhibited stronger model-based contributions to their choice behavior, suggesting that they utilized a model of the environment to prospectively evaluate choices. On the other hand, participants who generalized on the basis of surface similarity in the patterning task were more likely to show a response pattern characteristic of a pure model-free choice strategy, with little influence of a model-based strategy. However, the degree of feature-based transfer did not predict sensitivity to the previous trial's reward. It is somewhat surprising that these purportedly associative processes were not strongly related. However, *selective* effects of higher-order processes on model-based contributions to choice have been demonstrated previously (Otto et al. 2015). The finding that application of an abstract rule selectively predicts model-based choice suggests that both reflect sophisticated, resource dependent processes. On the other hand, feature-based transfer and model-free choice are generally characterized as reflexive and stimulus driven. Thus, differences in task requirements, stimuli and outcomes may have a greater impact on the expression of these processes, such that associations between them may be less clear, despite the possibility that they are served by a common system.

One interesting aspect of the data is that there was no relationship between the ability to verbalize the patterning rule, and use of rule-based transfer in the patterning task. Thus, there were a number of participants who were able to extract an abstract rule-structure from the task, but did not apply this to the novel transfer stimuli, and instead relied on a similarity-based process. This finding is consistent with the idea that the use of rule-based processes requires a level of behavioral flexibility and cognitive control, in order to overcome habitual or stimulus-driven responses when planning and executing action, which is directly in-line with recent data connecting cognitive control abilities to model-based choice (Otto et al., 2015). However, as rule-discovery itself is often an effortful process, more research is needed to understand how and when learning vs. applying rules relies on cognitive control. Likewise either uniting or distinguishing feature-based generalization from model-free choice requires further research. The approach we are advancing here, that is, characterizing what kinds of individual performance is stable across tasks, will be critical in answering these questions.

Acknowledgments

This paper was supported by an Australian Postgraduate Award to HJD, an Australian Research Council Discovery Grant DP150104267 to MBG and EJL, and a Visiting International Collaborator Support grant from the University of Sydney awarded to ARO.

References

Balleine B. W., & O'Doherty J. P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35*, 48–69.

Daw, N. D., Niv, Y., & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*, 1704–1711.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., Dolan R. J. (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69,* 1204–1215.

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

Harris, J. A., & Livesey, E. J. (2008). Comparing Patterning and Biconditional Discriminations in Humans. *Journal of Experimental Psychology: Animal Behavior Processes, 34*, 144-154.

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Kahneman, D. (2011). Thinking, Fast and Slow, First Edition (New York: Farrar, Straus and Giroux).

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143*, 668–693.

McLaren I. P. L., Forrest C. L., McLaren R. P., Jones F. W., Aitken M. R., Mackintosh N. J. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory, 108,* 185–195.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183–246.

Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science, 24*, 751–761

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-Memory Capacity Protects Model-Based Decision-Making from Stress. *Proceedings of the National Academy of Sciences, 110,* 20941-20946.

Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive Control Predicts Use of Model-based Reinforcement Learning. *Journal of Cognitive Neuroscience, 27,* 319-333

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

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

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*, 168–176.

Wills, A. J., Graham, S., Koh, Z., McLaren, I. P. L., & 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, 37*, 308 –316.

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