When Should We Expect Indirect Effects in Human Contingency Learning?

Daniel A. Sternberg (sternberg@stanford.edu) and James L. McClelland (mcclelland@stanford.edu)
Department of Psychology, Stanford University
Stanford, CA 94305 USA

Abstract

How do we learn causal relations between events from experience? Many have argued for an associative account inspired by animal conditioning models, but there is a growing literature arguing that *indirect effect*s in contingency learning depend on explicit cognitive processes. Our experiments explore the basis of two such effects: blocking and screening off. In Experiment 1, we gave participants an untimed explicit prediction task to replicate standard findings in the contingency learning literature in a novel domain. We obtained robust indirect effects when participants had a causal framework to constrain their reasoning. In Experiment 2, we reduced the time available for explicit recollection by reconstructing the task as a fast-paced RT task. Participants continued to show robust learning of direct relationships, as measured by response times, but there were no indirect effects. Experiment 3 followed up on whether participants in our RT task would produce indirect effects through explicit processes when given an opportunity to make a more deliberative prediction at test.

Keywords: Learning; causal reasoning; implicit learning

Introduction

A child goes out to dinner with his family and at the end of the meal experiences a strong allergic reaction. Upon discussion with the restaurant manager, the child's parents learn that the sauce for his entrée contained shrimp, and peanuts were used in his dessert. Suppose the child has never had shrimp before. If he has had a history of peanut allergies, one may be inclined to attribute the allergy to the peanuts; if he had never had a peanut allergy before, one may be more inclined to suspect an allergy to the shrimp.

We can consider the child's previous experience with peanuts as the *direct evidence* about whether peanuts cause an allergic reaction. This evidence, together with the shrimp-and-peanuts event, provides *indirect evidence* about whether shrimp causes one. If peanuts had previously caused an allergy, this tends to *block* the inference that shrimp causes one; if peanuts had not previously caused an allergy, this tends to *screen off* the shrimp – *increasing* the likelihood of this inference. Comparing the two cases, the scenario above describes a *direct effect* whereby the strength of the perceived causal relation between peanuts and allergy should be higher for the blocking pair compared to the screening pair. It also describes an *indirect effect* whereby the strength for shrimp will be higher in the screening pair compared to the blocking pair. Table 1 encapsulates this information.

Effects similar to the indirect effect described above have often been demonstrated in contingency learning experiments. In these experiments, participants see a number of pairings of cues and outcomes during training. At test, they are asked to rate the various cues' causal strengths or to make predictions about the likely outcomes for each cue. Early contingency learning researchers such as Alloy and Abramson (1979) and Dickinson and colleagues (1984) compared their findings to models of animal conditioning that automatically generate indirect effects (e.g., Rescorla & Wagner, 1972; Pearce & Hall, 1980). Indeed, a large class of error-correcting learning algorithms predicts these effects (Rosenblatt, 1958; Rumelhart et al., 1986; Sutton, 1988). Recent dual process models of implicit and explicit learning have employed error-correcting learning algorithms in the implicit component of the models (e.g., Ashby et al., 1998; Sun et al., 2005) – suggesting that indirect effects should be a basic outcome of an implicit learning system.

Table 1: An example of direct and indirect effects in a contingency learning paradigm.

Training	Single item	Pair
Blocking pair	B_1+	B_1B_2+
Screening pair	S_1-	S_1S_2+
Direct effect	**$B_1 > S_1$**	
Indirect effect	**$S_2 > B_2$**	

Complicating the error-driven account have been findings of retrospective effects like backward blocking (Shanks, 1985), where the order of compound and single item events are reversed (e.g., shrimp and peanuts before peanuts alone). These models do not directly predict retrospective effects. Various modifications to the error-correcting learning algorithm have been proposed to accommodate retrospective effects (Van Hamme & Wasserman, 1994; Dickinson & Burke, 1996). These models continue to predict indirect effects as a basic outcome of the learning process. Another approach has been to argue that retrospective effects are instead driven by the explicit retrieval of memories for previously experienced events (McClelland & Thompson, 2008).

More troubling are recent findings that suggest indirect effects are often quite fragile in contingency learning tasks. De Houwer and Beckers (2003) found that blocking was attenuated when participants were given a relatively difficult secondary task (discriminating between a high and low tone) during training and test phases. "High-level" constraints such as assumptions that cues are additive in their effects also appear to modulate the size of indirect effects (Lovibond et al, 2003; Beckers et al., 2005; cf. Livesey & Boakes, 2004). These findings have led some to argue that an explicit propositional reasoning

process drives indirect effects, and some have gone so far as to argue that all associative learning is best explained by a propositional account (Mitchell et al., in press).

Contingency learning experiments have a number of common attributes and it should not be taken for granted that their results would generalize if these attributes were altered. The experimental context is often one that allows nearly unlimited time to deploy explicit processing. These studies also normally use stimuli drawn from familiar domains such as foods and allergies or symptoms and diseases. It seems likely that participants bring to these tasks a great deal of prior knowledge about the kinds of causal relations that are likely to apply in the domain. In other cases, where researchers have employed what are arguably novel domains, participants are explicitly given a particular causal framing, e.g., "blickets are objects that make the machine go" (Sobel et al., 2004), and may also receive an explicit demonstration that if either of two objects is a blicket, the machine will go.

To begin a systematic exploration of some of the issues raised above, our experiments explore an identical set of contingencies in two very different task settings. In one setting (Experiment 1), participants are asked to make an explicit prediction of whether an outcome will occur given the current set of items. In the other setting (Experiment 2), items occur, followed very shortly in some cases by outcomes, and participants must respond very quickly when the outcome occurs; so quickly that they must rely on perhaps implicit expectations to anticipate the outcome in order to respond quickly enough. The former setting is the kind in which one may expect explicit reasoning processes to occur, whereas the latter shares features with many tasks thought to involve implicit learning (Lewicki et al., 1987; Cleeremans & McClelland, 1990). Our study therefore provides a direct basis for comparing the outcome of learning in these two very different kinds of task situations. Crossed with the task manipulation, we employed a framing manipulation: Half of the participants in each task were given an explicit causal framing similar to those used in experiments with the blicket detector (Sobel et al, 2004), and half were not given such a framing. Overall the experiments allowed us to compare direct and indirect contingency learning in two very different task settings, and to examine the importance of framing for both kinds of learning.

Experiment 1: Untimed Prediction

Method

Participants 50 members of a paid participant pool at Stanford University participated in this experiment. 4 were removed due to performance falling more than two standard deviations below the mean, resulting in a total of 46 participants (23 in each condition).

Design and Procedure The experiment consisted of two training blocks followed by a test block. Each training trial began with a box in the center of the screen, in which one or two objects appeared following a random delay of 1-3s between each trial. The object images were taken from *http://www.openclipart.org*. Participants were instructed to respond by pressing a key to denote their prediction of whether a dot was to appear on that trial and were given unlimited time to make their response. After making the prediction, they were shown the outcome. On dot outcome (+) trials, a green dot appeared to either the left or the right side of the box for 500 ms (the side was randomly chosen for each dot trial). On no-dot (-) trials, the screen did not change during the 500 ms outcome period. After the outcome period, participants were given both visual and auditory feedback. The visual feedback consisted of a point score that appeared in the center of the box. Correct predictions were followed by "10" appearing in green, while incorrect predictions were followed by "-5" in red. Participants were instructed that every 5 points was equal to $0.01 in payment. Auditory feedback consisted of a pleasant sound for correct predictions and a buzzer for incorrect predictions. Figure 1 shows a single training trial in the experiment. Training consisted of 264 trials (24 exposures to each training event). During the test phase, no feedback was given, and each event was shown 5 times, for a total of 60 test trials.

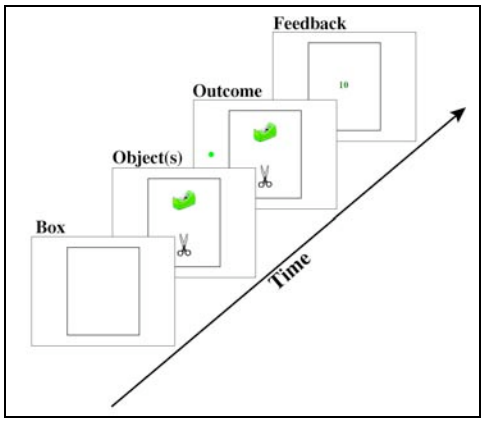

Figure 1: Layout of a single trial in the experiments

The structure of events shown to participants during training and test for Experiment 1 is given in Table 2. 11 events were shown during training, and 12 events were shown at test. Events that were likely to be followed by the dot ("+" events in Table 2) were followed by the dot on 22 of 24 presentations during training, versus 2 of 24 presentations for the other items. Object images were randomly matched to specific items for each participant. Filler events were included during training to lower the proportion of all trials in which the dot appeared, as pilot testing without these revealed that participants had a general bias to predict the dot outcome.

All participants received the same standard task instructions explaining the nature of the task and the payoff structure. Half of the participants (*framed* participants) in the study were given an additional set of

instructions immediately after the task instructions and immediately before beginning the experiment (shown in Table 3). These instructions explicitly indicated that some objects had the power to produce the dot while others did not, similar to the framing given to participants in blicket detector experiments (e.g. Sobel et al., 2004). The remaining (unframed) participants received no further instructions.

Table 2: Structure of events in Experiment 1

	Training		Test	
Tested items	Pair	Sing.	Pair	Sing.
Blocking	B_1B_2+	B_1+	B_1B_2	B_1 B_2
Screening	S_1S_2+	S_1-	S_1S_2	S_1 S_2
Control	C_1C_2+		C_1C_2	C_1 C_2
Negative	N_1N_2-	N_1-	N_1N_2	N_1 N_2
Filler items				
+ Singleton	PS+			
- Singleton 1		NS_1-		
- Singleton 2		NS_2-		
- Pair	NP_1NP_2-			

Table 3: Instructions for the framing manipulation.

The following information may help to improve your performance during the experiment.
Some of the objects that you will see during the experiment have the power to make the dot appear, while others do not.
The computer has randomly decided which objects have this power at the beginning of the experiment.
You cannot tell just by looking at the object whether it has the power to make the dot appear.
However, you can learn which objects have this power based on whether or not the dot appears when the object is inside the box.
If two objects appear inside the box and at least one of them has this power, the dot will usually appear.
Sometimes the box may malfunction, and the dot may occasionally fail to appear when it should, and may occasionally appear when it shouldn't.
If you can determine which objects have the power to make the dot appear, this will help you to make predictions during the experiment.

After the test phase, participants were given a number of additional tasks. First they were asked, *"How likely is it that the dot would occur if this object appeared by itself?"* and responded on a scale from 1 and 10. They were then given a forced-choice pair identification task, in which each test item appeared in the center of the screen, with the four respective training items below. They then performed the same likelihood rating task as before with the pair events. Finally, framed participants were asked to give ratings of causal strength for each test and training item. They were asked *"How likely is it that this object had the power to make the dot appear?"* and told to choose from six options: "Sure No", "Probably No", "Guess No", up to "Sure Yes". Unframed participants were not asked to make causal strength judgments.

Results

Figure 2 plots dot prediction rates for key direct and indirect items during the test phase. We compared the difference between blocking and screening training items (B_1 and S_1) as an index of direct effects, and the difference between the respective test items (B_2 and S_2) as an index of indirect effects. Likelihood ratings and causal strength ratings elicited the same pattern of significant and non-significant findings as predictions, and are not shown. Both conditions showed robust direct effects (B_1 vs. S_1) across all three measures using Wilcoxon's signed-rank tests, all $p < .001$. Only participants in the framed condition showed significant indirect effects (B_2 vs. S_2), which were significant across all measures, $p < .01$. All of these measures failed to show significant indirect effects for the unframed condition, all $p > .1$.

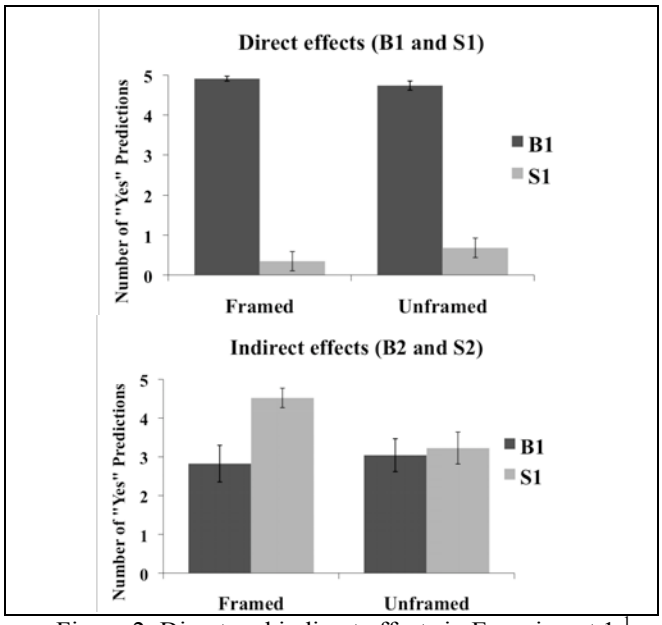

Figure 2: Direct and indirect effects in Experiment 1.[1]

Discussion

Experiment 1 demonstrated that our task could replicate findings in standard contingency learning tasks. By giving participants a clear causal framing of the experiment, we were able to obtain indirect effects, but there was no evidence of these effects in the absence of such a framing. It appears that such framing, either through explicit instruction or through stimuli from a well-learned domain, may be crucial to obtaining indirect effects in contingency learning experiments.

It is still possible that a faster-paced task in which subjects cannot explicitly recall prior events in order to generate a prediction might uncover an implicit learning process in which indirect effects may occur. Implicit learning has often been modeled using error-correcting learning models driven automatically by differences between predicted and observed outcomes (e.g., Sun et al,

[1] Means and standard errors are presented here for convenience. All statistical tests for prediction counts and ratings are non-parametric Wilcoxon's signed-rank tests. The prediction counts in particular are highly bimodal, with most participants either predicting the dot on all five of the test exposures or on none.

2005; Cleeremans & McClelland, 1990). With interleaved presentation of the blocking and screening training stimuli (B_1+, B_1B_2+; S_1-, S_1S_2+) such models produce indirect effects. On this basis, a theory in which implicit learning was based on such error-correcting learning would predict that we would obtain indirect effects in this experiment, regardless of whether participants were given an explicit causal framing.

Experiment 2: Fast-paced RT Task

In Experiment 2, we set out to modify our task to reduce the influence of explicit processing during both training and testing. We embedded the same event structure from Experiment 1 in a fast-paced serial reaction time task where participants were instructed to respond as quickly as possible to the appearance of the dot and to avoid responding when the dot failed to appear. If participants have learned the contingencies, they should respond faster to the dot when it follows the presentation of items that are associated with the dot outcome. Thus if we obtain a direct effect, response times for the B_1 item should be *faster* than for the S_1 item. If we obtain indirect effects, response times for the B_2 item should be *slower* than for the S_2 item.

Method

Participants 48 members of the Stanford Psychology paid pool participated in the experiment for payment. Three were removed because their performance during training fell more than two standard deviations below the mean. One participant was removed due to experimental error. This resulted in 22 participants in each condition.

Design and Procedure The overall structure of the training phase was identical to Experiment 1, except that there were twice as many training trials. In order to keep the contingencies as similar as possible subjects were given 43 dot outcome trials for each "+" item and 5 dot outcome trials for each "-" item. Participants completed 528 training trials comprising 48 exposures to each training item and 288 test trials, comprising 24 exposures to each test item.

Trials began similarly to Experiment 1. Each trial began with the box framing the center portion of the screen. Objects appeared after a random 1-3 second delay as in Experiment 1. However, in the current experiment, the outcome occurred automatically 350 ms after the object(s) appeared. Participants were instructed that they could earn points by pressing the spacebar within a brief time window after the dot appeared. The outcome period lasted for 500 ms, with a response deadline occurring somewhere within. The response deadline was initialized at 400 ms, and decreased at a constant rate every 10 trials during the first 200 training trials, then remained at 275 ms for the rest of the experiment.

Unlike in Experiment 1, test trials in Experiment 2 still included outcomes and feedback, and were indistinguishable from training trials in structure. As in Experiment 1, the test included the all the training items as well as the previously unseen test items. Each test event occurred 24 times. For the singleton test items, the contingencies established during training were not maintained. Instead, each singleton was followed by the dot on 12 of its 24 presentations. To help maintain the learned contingencies, pair events continued to hold the same contingencies as they had during training. At the conclusion of the test phase, participants were given the same set of concluding tasks as in Experiment 1.

Results

Figure 3 shows average response times for the direct and indirect items during the test block. Paired t-tests of the response times revealed that both groups responded faster to the dot on B_1 events compared to S_1, both $p < .001$. The direct effects also appeared in likelihood ratings and causal strength ratings using Wilcoxon's sign rank tests, all $p < .001$. There was no evidence of indirect effects in either condition for any measure, all $p > .3$.

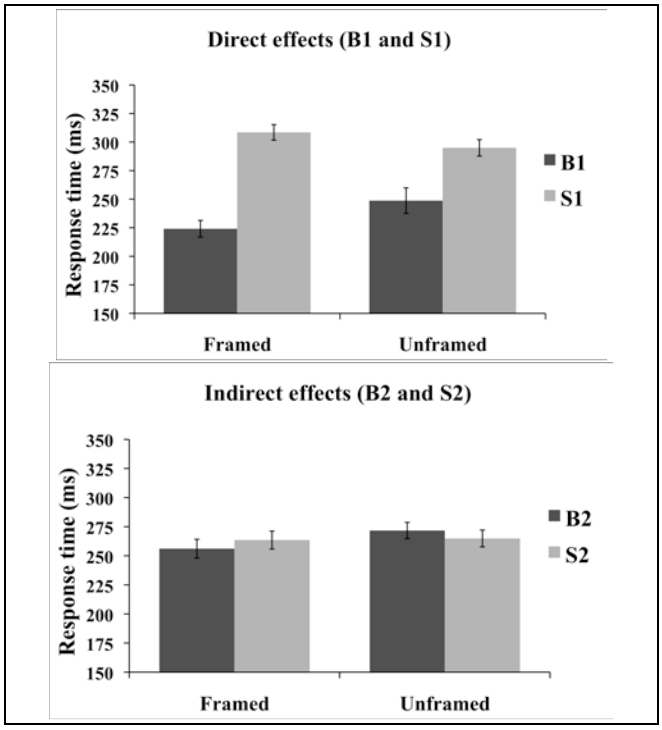

Figure 3: Direct and indirect effects as measured by response times in Experiment 2.

Discussion

In Experiment 2 we tried to minimize the influence of explicit processes during both training and test. If we were successful in uncovering an implicit error-correcting learning mechanism, we should have obtained indirect effects in both groups. Instead we found no evidence of any indirect effects. At the same time, participants continued to be respond in accordance with the direct contingencies that they experienced during training.

It also appears that participants do have explicit knowledge of the direct contingencies, as evidenced by their responses to rating and causal strength questions. While these explicit measures also failed to show any indirect effects, there is a possible confound. By breaking the contingencies for the singleton trials during the test phase, we may have reduced the strength of the contingencies that participants might have been able to use in answering the post-test outcome likelihood and causal strength questions. Experiment 3 provides a design that avoids this difficulty.

Experiment 3: Explicit Knowledge in the RT Task

Given our findings in Experiment 2, we wished to explore whether indirect effects in our tasks depend on time for deliberation during training or whether they could be generated at test given sufficient time to deliberate before responding. To address this question, we used a hybrid version of our task, training participants on the fast-paced RT task from Experiment 2, but placing them in the explicit prediction task from Experiment 1 during the test phase. This ensured that outcomes during the test phase would not wash out any of the effects over the course of test, and would maximize the likelihood of generating indirect effects if it was indeed possible to do so.

Method

Participants 25 members of the Stanford Psychology Department paid subject pool participated in the experiment for payment. Three participants were removed because their performance during training fell more than two standard deviations below the mean.

Design and Procedure The training phase was identical to Experiment 2. The test phase was identical to Experiment 1. Ratings and causal strength ratings were obtained after the test phase. All participants were given the framing instructions given to participants in the framed conditions of Experiments 1 and 2.

Results

Robust direct effects were observed for all three measures using Wilcoxon's sign rank tests, all $p < .001$. Blocking and screening test items only differed significantly in their causal strength ratings, $p = .046$, while the trend was non-significant for the other two measures, both $p > .2$. Figure 4 graphs the data from the power ratings.

Discussion

Our findings in Experiment 3 fall between those of Experiments 1 and 2. By giving participants the opportunity to make an explicit prediction immediately after training, we were still only able to obtain a significant indirect effect when we asked participants to rate how likely it was that each object "had the power to make the dot appear."

There are a few possible explanations for this finding. As participants showed weaker explicit learning of the direct contingencies in Experiment 3 compared to Experiment 1, there may have simply been a weaker overall indirect effect which happened to reach significance for the causal strength measure. Alternatively, there is some evidence that questions about cues' causal strengths generate more robust indirect effects than questions about predicting outcomes (Vadillo & Matute, 1998). It is also possible that the forced-choice pair identification task and pair ratings, which occurred immediately before the causal strength ratings, may have reinforced participants' memories of the events they had seen, increasing their ability to exploit these memories to the point of producing an indirect effect in some participants.

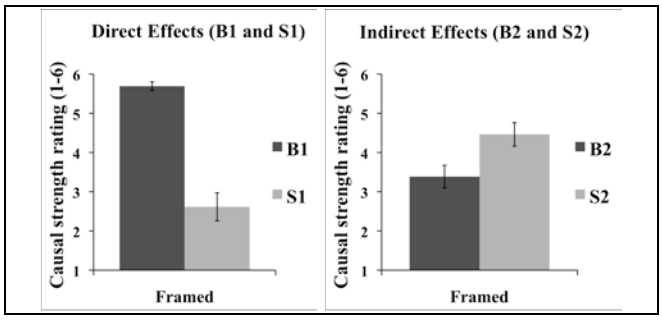

Figure 4: Causal strength ratings for direct and indirect items in Experiment 3.

General Discussion

In an explicit task, we found evidence that participants only learned indirect contingent relations when given a set of framing instructions explicitly providing a causal interpretation of the relationship between presented items and outcomes. It should be noted, however, that the use of so many different items may have influenced unframed participants' ability to use information from paired items that they might otherwise have exploited. We also note that framed participants were told explicitly that there would be occasional malfunctions, thus allowing occasional events not consistent with a simple causal story to be disregarded. Thus it is clear that further research will be crucial in order to more fully delineate the conditions necessary for obtaining indirect effects. In any case, the results from Experiment 1 make it unlikely that automatic error-correcting learning, as in some connectionist models provides the mechanism underlying contingency learning in our explicit prediction task. These conclusions appear to be consistent with a great deal of the recent human contingency learning literature.

Perhaps more surprisingly, our RT task failed to uncover any indirect effects whatsoever. This appears to be at odds with predictions of several models of implicit contingency learning (Cleeremans & McClelland, 1990;

Sun et al, 2005). We might still think that some kind of implicit learning process is operating in our fast-paced task given the task's time constraints and the fact that participants' response times show clear direct effects. If this is the case, however, the implicit learning system appears not to be operating through error-correction. Instead, our RT experiment results seem to be more in agreement with a Hebbian-like model. Of course, our experiment is not the final word on this point: it is possible, for example that an error-correcting effect would emerge with even longer training. Given that error-correcting models are more powerful and continue to be popular for simulating learning in a wide variety of domains such as language and categorization as well as animal conditioning, this is an important issue for further investigation. Further experimental and computational work is needed to elucidate the processes involved in both kinds of tasks explored in our investigations.

With these studies in hand along with other recent demonstrations of the fragility of indirect effects, we appear to be approaching a crossroads in the characterization of associative learning. There is now evidence that some effects formerly thought to arise as basic outcomes of an automatic associative learning process may require more complex processes than many had previously expected. It also seems that models of implicit learning will require further elaboration before they easily encompass the results of our RT experiment.

Acknowledgments

We wish to thank members of the PDP Lab at Stanford for helpful input on the work presented in this paper.

References

Alloy, L.B., and Abramson, L.Y. (1979). Judgement of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*, 441-485.

Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., and Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *305*, 442-481.

Beckers, T., Miller, R.R., De Houwer, J., and Urushihara, K. (2005). Reasoning rats: Forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General*, *135*, 97-102.

Cleeremans, A., & McClelland, J.L. (1990). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235-253.

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

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, *B, 49*, 60-80.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgment of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, A, 36,* 29-50.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 523-530.

McClelland, J.L., and Thompson, R.M. (2008). Using domain-general principles to explain children's causal reasoning abilities. *Developmental Science*, *10*, 333-356.

Mitchell, C.J., De Houwer, J., and Lovibond, P.F. (in press). The propositional nature of human associative learning. *Behavioral and Brain Sciences*.

Pearce, J.M., and Hall, G. (1980). A model for Pavlovian learning. Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532-552.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A.H., & Prokasy, W.F. (Eds.), *Classical conditioning II: Current theory and research.* New York: Appleton-Century-Crofts.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage in the brain. *Psychological Review*, *65*, 386-408.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume1: Foundations,* MIT Press, Cambridge, MA.

Shanks, D.R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology, 37*, 1-21.

Sobel, D.M., Tenenbaum, J.B. and Gopnik A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28,* 304-333.

Sun, R., Slusarz, P., and Terry, C. (2005). The Interaction of the Explicit and the Implicit in Skill Learning: A Dual-Process Approach. *Psychological Review*, *112*, 159-192.

Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*, 9-44.

Vadillo, M.A., & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *Quarterly Journal of Experimental Psychology: B*, *60*, 433-447.

Van Hamme, L.J., & Wasserman, E.A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*, 127-151.