International Journal of Comparative Psychology, 2009, 22, 61-74.

Covariation, Structure and Generalization: Building Blocks of Causal Cognition

Robin A. Murphy and Esther Mondragón
University College London, United Kingdom

Victoria A. Murphy
University of Oxford, United Kingdom

Theories of causal cognition describe how animals code cognitive primitives such as causal strength, directionality of relations, and other variables that allow inferences on the effect of interventions on causal links. We argue that these primitives and importantly causal generalization can be studied within an animal learning framework. Causal maps and other Bayesian approaches provide a normative framework for studying causal cognition, and associative theory provides algorithms for computing the acquisition of data-driven causal knowledge.

> And though we must endeavour to render all our principles as universal as possible, by tracing up our experiments to the utmost, and explaining all effects from the simplest and fewest causes, it is still certain we cannot go beyond experience; and any hypothesis, that pretends to discover the ultimate original qualities of human nature, ought at first to be rejected as presumptuous and chimerical (Hume, Introduction, 1740/2002).

The psychology of causes and causation involves core behavioral and cognitive functions and retains a central position in Experimental Psychology. Perception (Michotte, 1946/1963), learning (Le Pelley, 2004), pattern or rule acquisition (Conway & Christiansen, 2001; Murphy, Mondragón, & Murphy, 2008), instrumental action (Buehner, 2006; Wasserman, Elek, Chatlosh, & Baker, 1993) and analogical thinking (Penn, Holyoak, & Povinelli, 2008)—even issues related to questions of free-will (Haggard, Clark, & Kalogeras, 2002; see also Blaisdell, 2008)—underlie causal understanding. Causal model theory (Waldmann & Holyoak, 1992), causal map theory (Gopnik et al., 2004) and propositional reasoning (Mitchell, De Houwer, & Lovibond, 2009) are recent attempts to explain causal learning in humans. Similar theories are being developed to explain animal conditioning (e.g., Beckers, Miller, De Houwer, & Urushihara, 2006; Courville, Daw, & Touretzky, 2006). Much of this work is described in terms that seem to go beyond simple Humean association. We review experiments touching on components of causal learning studied in the conditioning chamber; statistical learning, structure learning and generalization of structure.

It is worth beginning by discriminating between two very different types of causal learning that themselves may invoke different representations and learning processes. We will only address one of these two. Young (1995) provides a useful distinction between personal causal theories, ones developed by direct observation, from public causal theories that are acquired and elaborated via social or cultural transmission. We argue that much of animal behavior can be explained via the same associative learning processes that guide human personal causal theories. We do not deny a role for theory-driven causal thinking acquired via logical reasoning

Correspondence concerning this article should be addressed to Robin A. Murphy, Institute of Behavioural Neuroscience, Division of Psychology & Language Sciences, University College London, Gower St., London, WC1E 6BT, U.K. (robin.murphy@ucl.ac.uk).

and social or cultural transmission even in animals but simply believe that a different set of explanations may be required (although see Heyes, 2005).

Experiential cause - effect learning involves acquiring a representation of the temporally and spatially contiguous stimulus events encountered in the world. How does an animal learn that its behavior causes a food pellet to appear? How does it learn that a tone seems to cause or (perhaps only) predict the occurrence food (e.g., Blaisdell, Sawa, Leising, & Waldmann, 2006)? These questions have been the interest of animal Pavlovian and Instrumental conditioning research for the last 100 years with recent advances in causal map theory challenging simpler associative explanations.

Causal Map Theory (e.g., Gopnik et al., 2004; Pearl, 1996) like Spatial Map theory (O'Keefe & Nadel, 1978) are descendents of the more general cognitive map theory of Tolman (1948) who used maps as a metaphor for explaining goal-directed behavior and the flexible representations of causal knowledge. A causal map is a developed representation of the causal links between events. Rather than simply containing associated links, these representations code both causal power or strength and temporal directionality. In addition to explaining the acquisition of causal links the maps can be used to predict what effects an intervention may have on the causal links as well, deducing the consequences of novel pathways in a causal net. For example, a virologist may use correlation and directionality to discover which virus correlates with which disease. They are also able to intervene to change the course of a disease by manipulating the conditions that support the virus. However, an important aspect of causal thinking that is under specified in causal maps, but is a crucial aspect of causal cognition, is an explanation of how knowledge might generalize between domains. So the virologist who has a causal theory of germs is also able to transfer this causal knowledge to discover new germs and develop new vaccines or perhaps to postulate the existence of antibiotics. The ability to see analogies across the causal structure of different sets of stimuli is argued to be what makes human thinking unique (Penn & Povinelli, 2007). How might one characterize this type of complex causal knowledge for the rat in a conditioning chamber?

In this paper we consider some of the evidence for learning causal primitives. The first one is well characterized by animal models of conditioning - strength of contingency. We then discuss some further data on temporal order and structure and the generalization of causal structure to novel stimuli.

Strength of Contingency

The British empiricists (e.g., Hume, 1740/2002) provided an initial model for how researchers might think about how animals learn causal relations. They assumed, in *tabula rasa* tradition, that causal knowledge was acquired through the senses rather than innately given. Acquiring and then using knowledge about the causal relations between events in the natural environment requires using sensory information to encode the relevant spatial and temporal contiguity between events (e.g., Baker, Murphy, Vallée-Tourangeau, & Mehta, 2001; Mackintosh, 1978). From this perspective no true understanding of causal power is acquired, in fact causation is a human mental construction (Young, 1995). All that is observable to the behaving animal and useful for predicting future events are the local

contiguous relations between events. So, temporal contiguity between experiencing food smells and the nutritional consequences of ingesting food encourages an association between stimuli that allows an animal to find similar food in the future. In other words, it acts as if it has causal knowledge. The actual causal mechanisms behind volatile chemicals that excite nasal receptors and the role played by nutrition for physical health are cognitively irrelevant to the animal. Contiguity as a cue to causality has been the natural point of contact between human philosophical ruminating about causation and animal cognition (Dickinson, 2001).

A theory of behavior, based simply upon a single contiguous pairing between a smell and food, is insufficient to explain how repeated similar experiences over time come to drive this behavior. Research with more direct relevance to causation comes from studies exposing animals to similar stimuli that are paired with similar outcomes over time.

Pavlov (1927) showed this with dogs that gradually acquired conditioned salivary responses (CR) in anticipation of food (the unconditioned stimulus; US). Importantly, the acquisition of the CR followed repeated presentation of a very similar conditioned stimulus (CS) with a very similar US over time. Taking this further, Rescorla and Lolordo (1966) showed that conditioning is not simply a reflection of CS-US pairing experience, with greater numbers of pairings contributing to greater behavior. Pavlovian conditioning involves animals being sensitive to or computing the overall rates or probabilities of the US in presence and the absence of the CS (see also Rescorla, 1968; 1969). In several studies Rescorla and colleagues varied the relative likelihood of the occurrence of a US (e.g., Shock) during training sessions in which an originally neutral but soon to be CS (e.g., Tone) was presented. Conditioning to the CS as a predictor of shock was controlled by whether the rate of US occurrence in the presence of a CS was i) greater than, ii) less than or iii) equal to the rate of US occurrence in its absence. Learning occurred regardless of the number of pairings of the two events but the strength of conditioned responding was proportional to the strength of the contingency. These three contingencies fostered excitatory, inhibitory or no conditioning.

Conditioning to a CS occurs to the extent that it signals a change in US likelihood from that observed in the absence of the CS. On the surface this is much like Kelley's (1973) explanation of how humans make causal attributions,"An effect is attributed to the one of its possible causes with which, over time, it covaries" (Kelley, 1973 p. 108). Rescorla and colleagues' experiments (1966, 1968, 1969) showed that conditioned behavior required CSs that covaried with the US.

Although we have hinted at this point earlier, it is important for much of the rest of the discussion to clearly distinguish a computational theory of covariation learning from the algorithm that might explain how covariation is computed. Marr described this distinction quite succinctly in relation to vision (Marr, 1982). The visual system computes the relations required for 3-D perception without an internal 3D representation. This distinction in relation to learning has also been made (Baker, Murphy, & Vallée-Tourangeau, 1996; Cheng, 1997) but is worth repeating. An animal's behavior changes as the experimenter manipulates the covariation, if the animal's behavior is sensitive to these changes then the animal has a developed a representation of the covariation. Covariation is

computed by the experimenter in a number of possible ways; Pearson's moment correlation (*r*) provides a measure of how two variables covary. With one-way binary events Allan's (1980) $\triangle p$ statistic captures the relation. One (relatively) simple representation that accounts for the acquisition of this relation is a single associative connection. The associative connection is acquired via a competition between the target stimulus and the contextual cues for control over activation of the US representation. One algorithm that can compute this value is the Rescorla-Wagner model (Rescorla & Wagner, 1972; Chapman & Robbins, 1990). We have described the relation between these two levels of analysis in more detail elsewhere arguing that the Rescorla-Wagner model provides an associative explanation of the representations of covarying stimuli (Baker et al., 1996).

The Rescorla-Wagner model (Rescorla & Wagner, 1972) was designed to provide an associative account for contingency learning effects in animal learning. It describes how learning involves acquiring and updating associations formed between a stimulus and its outcome. In conditioning terms this is the association between the CS and the US. Associations get stronger to the extent that a discrepancy exists between the activation of the US representation (that the CS can cause) and the maximum level of activation caused by the US itself. If there is a difference between these two, learning will occur. The model also predicts that on any given trial the cause or predictor of the outcome is unknown and so all potential causes or predictors of the US share the associative strength actioned by the occurrence of the unpredicted US.

In Rescorla's experiments (e.g., 1968) the competition for association is hypothesized to be between the Tone (CS) and the cues of the context. Our research interest in this phenomenon has been to explore this prediction. A CS trained as a cause of a US will acquire positive associative strength at the expense of the context whereas a CS trained without such a relation occasions the context as an associative predictor of the US. We have conducted a number of studies designed to directly examine this reciprocal conditioning prediction with both rats and humans. We looked specifically for patterns of reciprocal learning as well as certain frequency learning effects predicted by the associative model.

In one set of appetitive Pavlovian conditioning experiments, rats were presented with Light-Food (i.e. CS-US) relations defined by the occurrence of a 10 second Light paired with food pellets (Murphy & Baker, 2004). Six different groups of rats received 40 trials. For half of the sample (3 groups) the light was positively related to the occurrence of food while for the remaining 3 groups the light was unrelated to food occurrence. The likelihood of food on a trial with the light (p(US|CS) was 50% greater than on trials when the light was not present (p(US|noCS), the difference between these two conditional probabilities for food is captured by a moderately positive relationship. Allan's (1980) one-way contingency between the CS and the US is expressed by the difference between these two conditional probabilities. In our case this was $\Delta p = 0.5$. The overall likelihood of food in the absence of the light also varied between the three groups (0%, 25% and 50%). The three groups of rats receiving the unrelated Light-Food training received food on either 25%, 50% or 75% of trials regardless of the presence of the light. For these three groups, the Light signaled no change in the likelihood of food ($\triangle p = 0.0$). Much like Rescorla's finding, rats showed statistically stronger nose-poke responding to the CS when it signaled an increased

likelihood of the US. Additionally, we found evidence that cues signalling a greater proportion of USs generated more responding regardless of the overall relation. Evidence of the reciprocal conditioning between the CS and the context came from an assessment of responding to a contextual cue, in our case a physical metal bar that entered the chamber. Learning about the context was measured by contact with it. In measuring responding to this cue we were able to demonstrate that as the predictiveness of the CS decreased, animals were attracted to and made contact with the context cue, as if it was the more reliable predictor of food. This result is consistent with the theoretical account Rescorla and Wagner provided for sensitivity to covariation (Rescorla & Wagner, 1972). Sensitivity to the covariation between cues is dependent upon the causal context within which the covariation is experienced.

One possible explanation for our results is based on response competition. Response competition explanations are a simple way for animals to behave as if they have internalized a representation of the causal relationship. Animals perhaps do not learn about the context when they are responding to the CS, or do not learn about the CS when they are responding to the context. This explanation still requires that the contingency directs animals to the correct response. Alternatively, Nose-pokes and context responding might be incompatible and therefore only one or the other response is possible. A response competition account of covariation sensitivity requires a relatively simple peripheral account of how the animals 'solve' the problem. Nevertheless, response competition may be the behavioral solution for the computation problem of contingency learning (see also Dwyer, Starns, & Honey, in press, for a similar explanation of causal interventions in Blaisdell et al., 2006). We extended the analysis of CS/Context interactions by studying humans in a causal discovery task that presumably had none of the response competition issues.

Research on human covariation detection has suggested that the same associative model accurately predicts human acquisition of covariation (e.g., Baker, Berbrier, & Vallée-Tourangeau, 1989; Jenkins & Ward, 1965; Wasserman et al., 1993). In one experiment from our lab designed to mirror the rat experiment we provided undergraduate students with a causal discovery task. Students were required to learn about various fictitious virus-disease relations. Much like the virologist of our original example, participants discovered the contingency between each virus and a specific disease. Similar to the CS-US (Light-Food) relation that we trained the rats, students were trained with Virus labels that preceded the disease effects. Causal judgments like rat nose-pokes reflected both the degree of contingency between the two events (Positive or Zero) and the frequency of the disease (Vallée-Tourangeau, Murphy, Drew, & Baker, 1998). The evidence for the reciprocal learning between stimulus and context, as a function of variations in contingency, points directly to the involvement of associative processes in the computation of covariation. As important as covariation is, elaborated causal map theory stresses the importance of other variables to support causal knowledge, namely the temporal or structural relations for causal learning.

Temporal Structure

Causal structure in human learning has been argued to be as important as causal strength for causal reasoning since one needs to know which event is the cause and which one the effect (Lagnado, Waldmann, Hagmayer, & Sloman, 2007). It is claimed that associative models like the Rescorla-Wagner model (RWM) do not take temporal order or structure into consideration (Griffiths & Tenenbaum, 2005; Waldmann & Holyoak, 1992). There are two senses in which the RWM addresses causal structure and temporal order. The first is that the model predicts acquisition between the predictor and the outcome. The order in which the events are experienced dictates which association is calculated. The association describes the ability of the CS to activate the US, and is not usually bidirectional. Trained with a set of 10 CS→US pairings the model predicts that the CS acquires a strong positive association with the US, but not a strong positive association between the US and the CS; in other words, presentation of the US will not excite the CS. Conversely, given 10 US→CS trials the model predicts that the CS acquires a strong negative association with the US. Positive associations indicate causal relations while negative associations indicate preventative relations. The empirical evidence for this effect in the conditioning literature is strong.

Moscovitch and Lolordo (1968) describe a set of experiments showing how the behavior acquired when a CS is presented before a US is markedly different from the behavior acquired when the CS follows the US. With standard forward pairing of the CS and US animals treat the CS as a predictor of shock occurrence. Trained with the same stimuli with the order of the stimuli reversed they treat the CS as a signal that the shock has already occurred. (see also Holder & Garcia, 1987).

Conditioned behavior therefore, is influenced by the ordering of the cues but this evidence does not tell us whether the temporal order of the cues is encoded during conditioning. Causal map theory implies that each causal event pair represents the strength of the association as well as the temporal structure (Lagnado et al., 2007). What is the evidence then that animals encode these relations? Encoding temporal order requires demonstrating that animals can discriminate A→B from B→A.

Seger and Scheur (1977) demonstrated that rats could discriminate trials with Tone→Light→Food from those with the order of the two CSs reversed Light→Tone→no Food. The confound for any interpretation that requires temporal order encoding is that the rats might have learnt to respond to the second part of the trial and simply solved the discrimination by responding to the late occurring Light (see also Weisman, Wasserman, Dodd, & Larew, 1980).

A series of appetitive Pavlovian conditioning experiments that unambiguously shows temporal order acquisition was performed in our lab. Rats were required to learn that the order in which two neutral cues were presented signaled whether food was going to be delivered, but where every cue was at the beginning and end of both reinforced and nonreinforced pairs (Murphy, Mondragón, Murphy, & Fouquet, 2004). In these experiments neutral auditory and visual cues (e.g., lights, tones, clicks) were trained in pairs such that each of four cues had four different roles. They were each the first and second cues of both rewarded and nonrewarded trials. In this way nothing about an individual stimulus

informed the rats of the possible outcome of the trial. Table 1 illustrates the design utilized in these experiments. Although a minimum of three stimuli are needed to answer this question, using four stimuli has the advantage that it allows that pairs of cues have a balanced role in terms of their signaling relation to the other two cues. So both A and C precede B and D but signal the reversed contingency between the cues and the US. We trained rats with 5-second exposures to individual elements and presented food pellets following half of the orders. The measure of conditioned behavior during the second stimulus of each pair was whether they would come to nose-poke the food tray in anticipation of food on the appropriate trials. Although a somewhat difficult discrimination for rats to learn they did come to solve this problem. Although temporal order is not represented, *per se*, as a variable in the RWM, other versions of associative models can quite easily incorporate this order (e.g., Baetu & Baker, in press). We discuss the application of associative principles to problems of temporal order in more detail in the following section.

Table 1
Experimental design used in Murphy et al. (2004) to study temporal order discrimination. Each stimulus (A, B, C and D) was trained as the first and second stimulus and was paired with reinforcement (Rf+) and nonreinforcement (nRf-).

Reinforcement conditions	Rf +	nRf-
Stimulus pairs	A→B	B→A
	B→C	C→B
	C→D	D→C
	D→A	A→D

In summary, the evidence is that temporal order, if not always coded during learning, can be acquired given the appropriate causal structure training. This is not a new suggestion, although these are new data on the extent to which temporal order can be acquired and used as a discriminative cue for behavior. Miller and colleagues have proposed that the entire temporal structure of training is acquired during conditioning. The Temporal Encoding Hypothesis suggests that rats code temporal information about the CS and its relation to the US (e.g., Barnet, Grahame, & Miller, 1993; Blaisdell, Denniston, & Miller, 1998) with a fine-grained temporal detail.

Rule Learning and Generalization

The evidence that rats can discriminate A→B from B→A may be argued to be at the very low end of complexity in causal map terms. Other more complex causal ordering experiments have been conducted with rats that indicate that rats can solve these causal problems. Blocking (Kamin, 1969), Backward blocking (Miller & Matute, 1996), and the difference between common cause and common effect training (Beckers, Miller, De Houwer, & Urushihara, 2006) have all been

tested. Generally, findings are consistent with a Bayesian Causal map analysis. However, this might not be surprising since this analysis provides a justification for many possible results depending upon the causal map that is acquired.

Complex serial ordered events are also important to causal map theory and have been studied in animals in relation to serial and pattern learning. The advantage that three element sequences, or higher, have over two-element pairs is that three elements allow the experimenter to present more complex causal relations among stimuli as a cue for behavior. Three elements also allow us to ask questions as to whether animals can extract the causal chains present in the experience. There is already a considerable body of evidence from studies of animal behavior that chains of sequences can be learnt (e.g., Capaldi & Miller, 2001; Fountain, 2008; Gentner, Fenn, Margoliash, & Nusbaum, 2006). In the experiment we describe here, we used pairs of stimuli to construct three element sequences to test for temporal structure as well as extraction of pattern learning in rats. direr

Prelinguistic infants (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996), primates (Hauser, Weiss, & Marcus, 2002) and pigeons (Hebranson & Shimp, 2003) can learn simple rules that describe sequential relations. The reason we liken these to causal maps is that they contain some of the important elements of a causal map (Gopnik et al., 2004). Causal relations are a form of rule which describes the temporal relation between chains of events. Although the experiments that form the basis for our study with rats were conducted ostensibly to study innate language abilities in infants they embody simple rules that have the characteristics of a causal relation (see also Katz, Wright, & Bodily, 2007; Penn & Povinelli, 2007).

These experiments used a modification of a pattern learning design originally suggested to explain early evidence of language learning in babies. Using a habituation procedure Marcus et al., (1999; see also Saffran et al., 1996) exposed 7-month-old infants to strings of phonemes that obeyed a rule (e.g., XYX). When subsequently exposed to novel phoneme strings that either did or did not adhere to the habituated rule, 15 out of 16 babies gazed longer in the direction of the sound source when the inconsistent sequences were played. These researchers took this finding to suggest that the babies had learnt a rule in the initial exposure phase and could therefore discriminate the habituated rule during the test phase. From the perspective of causal map theory this procedure is analogous to training a cyclic causal map in which events in a causal chain reoccur. The transfer phase to novel items is of some interest to researchers of causal cognition since the ability to transfer suggests that the learned rule was flexible and could be applied in novel domains, albeit within the same sensory modality.

We examined whether rats could learn to resolve three-element sequential rules (Murphy, Mondragón, & Murphy, 2008). In our first experiment hungry rats were exposed to sequences of stimuli (e.g., light and dark) like the two element sequences used to test learning of temporal order except now rats were exposed to three-element sequences (light-dark-light) that obeyed rules. Items consistent with the rule (e.g., A-B-A) were paired with food reinforcement while those that violated the rule were extinguished. The test of acquisition required examining nose-poking behavior on the third element of each trained sequence. Each element

was present on both reinforced and nonreinforced trials; the animals could only use the previous two cues to determine whether it was appropriate to anticipate food on a given trial. Rats were able to learn this discrimination suggesting acquisition of the temporal structure of these cues.

The design of the experiment allows at least two possible conclusions about how multiple instances of each rule were learnt. Animals may learn for example that ABA and BAB signaled food but do not perceive them as from the same category. The sequences may have been treated as behaviorally equivalent because of their consequences but the rats may not have an internal representation of the extracted rule. Evidence from research on acquired equivalence (Honey & Watt, 1998) suggests that animals can generalize between training stimuli that signal the same consequences. In our case both instances of our trained rules had the same consequences and they were behaviorally equivalent, providing evidence of sequential and rule learning in rats.

In a second experiment we tested whether rats would show the generalization of responding, that is would they treat novel stimuli that obeyed the same pattern as instances that were similar to the reinforced stimuli from training. As in the previous experiment rats were trained with three element auditory sequences comprised of two pure tones (A = 3.2 kHz, B = 9 kHz). Following training with the XYX rule rats were exposed to novel sequences that obeyed or did not obey the learned rule. These novel instances used auditory stimuli outside of the training range (C = 12 kHz, D = 17.5 kHz). The pairs of stimuli were counterbalanced for half the animals. During a test of the novel cues rats responded more to the novel sequences that obeyed the rule. This evidence suggests that temporal sequences are both learnable and transferable. The underlying abstract rule that relates the elements was transferred to the novel cues. We have been arguing that this sort of ability has something in common with a causal rule. This rule describes the temporal relation between events and an outcome, and is flexible enough to be used in novel domains. It is worth adding that in spite of behaving on the basis of the rule, the animals may not be acquiring anything like a symbolic representation of the rule.

We may ask how the rats might solve this problem. They could perhaps transpose a sort of tune from training to transfer stimuli. A simple associative account (Mackintosh, 1965; Spence, 1937) based on the generalization gradients generated by the specific stimuli could not explain the results. If on the transfer test the elements are transposed simply as a result of a generalization process in which the common elements are ignored and the unique elements acquire the critical positive or negative strength, there would be no basis for discrimination. All the instances of our training sequences both reinforced and nonreinforced, share the same common elements and the same unique elements (e.g., A and B). All the instances of the transfer sequences also share the same and unique stimuli (e.g., C and D). Thus, any common elements shared between training and transfer would be identical for all sequences. The only source for discrimination requires some sort of temporal encoding. Consequently, a simple generalization model of transposition can not explain these results. Several possible associative mechanisms are possible. For instance Wallace and Fountain (2002) developed a Sequential Pairwise Associative Memory (SPAM) to account for pattern of numbered stimuli (reward magnitude).

Recently Baetu and Baker (in press) have modeled three element chain sequences using a simple auto-associator. In their model all three stimuli in a sequence A→B→C become associated as well as context cues that are associated with pairs of stimuli. Temporal order was modeled by varying the order in which the elements of the stimulus chain were activated. With no *a priori* assumptions concerning the causal structure, the associative model settled on a set of weights that modeled specific aspects of the human data to which the model was compared. The humans were taught sequences of visual stimuli. Interestingly, specific aspects of causal map theory, such as adherence to the Markov rules, were obeyed by the model and the human participants. Importantly from an associative perspective, the acquisition of a causal map was performed with no assumptions and only the basic principles of association.

One aspect of causal thinking that these models cannot account for is the transfer data. We believe that the addition of an associative generalization mechanism between elements in the sequence may allow these associative models to predict the pattern of generalization findings of our final experiment and we are currently investigating this idea. A similar idea has been used by Haselgrove, George, and Pearce (2005) to understand generalization of spatial structure.

Penn and Povinelli (2007) have argued that what sets animals and humans apart is the ability to analogize complex causal relations. Complex in that they involve sequences of interrelated directional relationships. At the conference in Belgium that initiated this special issue, Derek Penn presented the interesting example of how causal relations are used to foster analogies and causal discovery. The scientific model developed to understand the forces and relations between celestial bodies in the solar system provided a useful tool for predicted the movement of light in the night sky. This causal framework involves directional relations and forces describing a number of physical objects. The model provided scientists with novel predictions and more generally provided an important milestone in human understanding of the cosmos. He went on to argue that the ability to use this causal model in novel domains such as in the development of a theory of the structure of the atom was an example of the causal analogizing process (putting aside the problem that the analogy was fundamentally incorrect). Surely, a rat has no corresponding ability. But we can ask whether rats have the ability to learn sequences of directionally specific related events? Further can they generalize this knowledge to novel stimuli that have never been experienced? The usefulness of a causal model is in its ability to be transferred to novel contexts. Causal relations are rules that describe how stimuli relate to one another and are based at least partially on empirical evidence on how events are contingently related. Temporal order, sequence learning and a process of generalization might explain how causal maps are formed.

Causal map theory provides important value for the computational understanding of the nature of causal problems. We argue these theories are less helpful in understanding how the brains of animals (including humans) come to solve causal problems (e.g., Baker, Baetu, & Murphy, in press; Baker et al., 1996). Normative descriptions of causal thinking provide useful computational analyses of causal problems but may not explain the brain's algorithms for extracting causal relations. The analogy between humans' and animals' causal cognition may be closer than we think.

References

Allan, L. J. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*, 147-149.

Baetu, I. & Baker, A. G. (in press). Human judgments of positive and negative causal chains. *Journal of Experimental Psychology: Animal Behavior Processes*.

Baker, A. G., Baetu, I., & Murphy, R. A. (in press). Propositional learning is a useful research heuristic but it is not a theoretical algorithm. *Behavioral and Brain Sciences*.

Baker, A. G., Berbrier, M. W., & Vallée-Tourangeau, F. (1989). Judgments of a 2x2 contingency table: Sequential processing and the learning curve. *Quarterly Journal of Experimental Psychology, 41B*, 65-97.

Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.),*The Psychology of Learning and Motivation* (Vol. 34, pp. 1-45). San Diego: Academic Press.

Baker, A. G., Murphy, R. A.,Vallée-Tourangeau, F., & Mehta, R. (2001). Contingency learning and causal reasoning. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of Contemporary Learning Theories*. (pp. 255-306) Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Barnet, R. C., Grahame, N. J., & Miller, R. R. (1993). Local time horizons in Pavlovian learning. *Journal of Experimental Psychology: Animal Behavior Processes, 19*, 215-230.

Beckers, T., Miller, R. R., De Houwer, J., & Urushihara, K. (2006). Reasoning rats: forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General, 135*, 92-102.

Blaisdell, A. P. (2008). Cognitive dimension of operant learning. (pp 173-195). In H. L. Roediger, III (Ed.), *Cognitive Psychology of Memory. Vol. 1 of Learning and Memory: A Comprehensive Reference*, 4 vols. (J. Byrne Editor). Oxford: Elsevier.

Blaisdell, A. P., Denniston, J. C., & Miller, R. R. (1998). Temporal encoding as a determinant of overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 72-83.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, 311, 1020-1022.

Buehner, M. J. (2006). A causal power approach to learning with rates. In R. Sun & N. Miyake (Eds.), Proceedings of the 28th Annual Conference of the Cognitive Science Society. Mahwah, New Jersey: Erlbaum.

Capaldi, E. J. & Miller, R. M. (2001). Molar vs molecular approaches to reward schedule and serial learning phenomena. *Learning and Motivation, 32*, 22-35.

Chapman G. B. & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition, 18*, 537-545.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367-405.

Conway, C. M. & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5, 539-546.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Science, 10*, 294-300.

Dickinson, A. (2001). The 28th Bartlett Memorial Lecture. Causal learning: an associative analysis. *Quarterly Journal of Experimental Psychology, 54B*, 3-25.

Dwyer, D. M., Starns, J., & Honey, R. C. (in press). Causal reasoning in rats: A re-appraisal. *Journal of Experimental Psychology: Animal Behavior Processes*.

Fountain, S. B. (2008). Pattern structure and rule induction in sequential learning. *Comparative Cognition & Behavior Reviews*, 3, 66-85.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204-1207.

Gopnik, A., Glymour, C., Sobel, D. M., Shulz, L., Kushmir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 3-32.

Griffiths, T. L. & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334-384.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*, 382-385.

Haselgrove, M., George, D. N. & Pearce, J. (2005). The discrimination of structure: III. Representation of spatial relationships. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 433-448.

Hauser, M. D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition, 86*, 15-22.

Hebranson, W. T. & Shimp, C. P. (2003). "Artificial grammar learning" in pigeons: A preliminary analysis. *Learning & Behavior, 31*, 98-106.

Heyes, C. (2005). Imitation by association. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Mirror Neurons to Memes*. Cambridge, MA: MIT press.

Holder, M. D. & Garcia, J. (1987). Role of temporal order and odor intensity in taste-potentiated odor aversion. *Behavioral Neuroscience, 101*, 158-163.

Honey, R. C. & Watt, A. (1998). Acquired relational equivalence: Implications for the nature of associative. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 325-334.

Hume, D. (1740/2002). A treatise on human nature. Guttenberg etext. http://www.gutenberg.org/dirs/etext03/trthn10.txt

Jenkins, H. & Ward, W. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs, 7*, 1–17.

Kamin, L. J. (1969). Selective association and conditioning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental Issues in Associative Learning*. Halifax, NS: Dalhousie University Press.

Katz, J. S., Wright, A., & Bodily, K. D. (2007). Issues in the comparative cognition of abstract-concept learning. *Comparative Cognition & Behavior Reviews, 2*, 79-92.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107-128.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology Philosophy and Computation*. Oxford: Oxford University Press.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology, 57B*, 193-243

Mackintosh, N. J. (1965) Transposition after 'single-stimulus' training. *The American Journal of Psychology*, 78, 116-119.

Mackintosh, N. J. (1978). Cognitive or associative theories of conditioning: Implications of an analysis of blocking. In S. Hulse, H. Fowler, & W. K. Honig (Eds.) *Cognitive Processes in Animal Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month old infants. *Science, 283*, 77-80.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Michotte, A. E. (1946/1963). *The Perception of Causality*. London: Methuen & Co.

Miller, R. R. & Matute, H. (1996). Animal analogues of causal judgment. In D. R. Shanks,

K. J. Holyoak, & D. L. Medin (Eds.), *The Psychology of Learning and Motivation* (Vol. 34, pp. 1-45). San Diego: Academic Press.

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

Moscovitch, A. & Lolordo, V. M. (1968). Role of safety in the Pavlovian backward fear conditioning procedure. *Journal of Comparative and Physiological Psychology, 66*, 673-678.

Murphy, R. A. & Baker, A. G. (2004). A role for CS-US contingency in Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 30, 229-239.

Murphy, R. A., Mondragón, E., Murphy, V. A. (2008). Rule learning in rats, *Science*, *319*, 1849-1851.

Murphy, R. A., Mondragón, E., Murphy, V. A., & Fouquet, N. (2004). Temporal order of CSs as a discriminative cue for conditioned responding. *Behavioral Processes, 67*, 303-311.

O'Keefe, J. & Nadel, L. (1978). The hippocampus as a cognitive map. Oxford: Oxford University Press.

Pavlov, I. (1927). *The conditioned reflexes*. London: Dover Press.

Pearl, J. (1996). Structural and probabilistic causality. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.),*The Psychology of Learning and Motivation* (Vol. 34, pp. 393-435). San Diego: Academic Press.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109-130.

Penn, D. C. & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology, 58*, 97-118.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66*, 1-5.

Rescorla, R. A. (1969). Conditioned inhibition of fear resulting from negative CS-US contingencies. *Journal of Comparative and Physiological Psychology, 67*, 504-509.Rescorla, R. A. & Lolordo, V. M. (1965). Inhibition of avoidance behavior. *Journal of Comparative and Physiological Psychology, 59*, 406-412.

Rescorla, R. & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. Black & W. Prokasy (Eds.), *Classical Conditioning II: Theory and Research*. New York: Appleton Century Crofts.

Seger, K. A. & Scheuer, C. (1977). The informational properties of S1, S2 and the S1-S2 sequence on conditioned suppression. *Animal Learning & Behavior, 5*, 39-41.

Saffran, J., Aslin, E., & Newport, E. (1996). Statistical learning by 8 month old infants. *Science, 274*, 1926-1928.

Spence, K. W. (1937).The differential response in animals to stimuli varying within a single dimension. *Psychological Review, 44*, 430-444.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*, 189-208.

Vallée-Tourangeau, F., Murphy, R. A., Drew, S. & Baker, A. G. (1998). Judging the importance of constant and variable candidate causes: A test of the Power PC theory. *Quarterly Journal of Experimental Psychology, 51A*, 65-84.

Waldmann, M. R. & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222-236.

Wallace, D. G. & Fountain, S. B. (2002). What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 43-63.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 19,* 174-188.

Weisman, R. G., Wasserman, E. A., Dodd, P. W. D., & Larew, M. B. (1980). Representation and retention of two-event sequences in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 6*, 312-325.

Young, M. (1995). On the origin of personal causal theories. *Psychonomic Bulletin & Review*, *2*, 83-104.