Rational Rats: Causal Inference and Representation

Aaron P. Blaisdell

University of California, Los Angeles

Michael R. Waldmann

University of Göttingen

Word Count: 11290

Introduction

"When we try to pick out anything by itself, we find it hitched to everything else in the Universe." –John Muir (*My First Summer in the Sierra*, Boston: Houghton Mifflin, 1911)

There are three types of questions we use to interrogate our world: “What?” and “Who?” questions are about the objects and agents that populate our world; “Where?” and “When?” questions are about the spatial and temporal locations of those objects; “Why?” and “How?” questions are concerned with the causality of our world. Although these questions are all in the purview of our epistemological quest, questions about causality have remained the most intractable, both scientifically and philosophically.

Modeling the universe is a daunting task given its sheer complexity; and, hidden variables abound. Yet, modeling pieces of the universe at extremely local scales has become a straightforward and tractable endeavor. With a simple set of starting principles, we can isolate and dissect a system of interacting forces (by constructing what is called a free-body diagram), such as a wheel and axel, a mortar and pestle, a cue stick and a set of billiard balls, or a squirrel clinging to a tree branch. Using the cumulative knowledge of science and engineering, we can understand more complex causal systems, such as aplysia neural networks and mouse genomes; we can even create relatively complex fully functional (ideally) causal systems, such as an automobile engine or a computer and its software.

It perhaps comes as no surprise that philosophy has recognized the role of causality as the “cement of the universe” (Mackie, 1974) that underlies the orderly relations between observable events. Psychologists have also acknowledged that “to be a successful agent, we need to have causal representations which mirror the causal texture of the world" (Blaisdell, 2008; Tolman & Brunswik, 1935; Waldmann, Cheng, Hagmayer, & Blaisdell, 2008).

Causal Learning in Humans

Humans are causal agents par excellence. But, what are the psychological processes that have evolved to produce human causal cognition? And, which aspects of causal cognition are uniquely human and which are shared with other species? In this chapter, we describe how rational models of causal cognition can usefully frame these questions and allow us to design experiments which can illuminate the underlying psychological processes. We use a rational model as a starting assumption in our experiments. We start with the assumption that organisms such as rats and people are rational in that they have been designed by evolution to make decisions in an optimal manner based on information they have acquired through experience. This is just an assumption of convenience and it is used as a heuristic. We do not mean to imply that rats or people ARE rational, only that a rational model makes testable predictions about their behavior under certain conditions.

In this chapter, we discuss investigations of rat behavior under conditions designed to test the predictions of our rational model. If we find their behavior to be consistent with the rational model, then we do not take this to mean that we were right in that rats are rational. Our very next experiment could still falsify this assumption. This assumption should be tested repeatedly and systematically to prevent us from becoming dogmatic and blindly accepting this assumption. Here, we show how the rationality (i.e., optimality by design) assumption has proved useful in uncovering interesting and unanticipated behaviors in the rat.

Many of the events of the world appear to us to be directly connected by cause-effect relationships. The philosopher David Hume questioned this view in his seminal writings (e.g., Hume, 1748/1977). Hume looked at situations in which he observed causal relations and did not detect any empirical features which might correspond to evidence for causal powers. Causal power is the intuitive notion that one thing causes another by virtue of a hidden, unobserved power or energy that it exerts over the other (see Dowe, 2000). That is, causal power involves the inference of the transference of force, energy, or a conserved quantity such as between two colliding billiard balls or the change in charge of a photocell when a photon collides with it. Hume did not find any evidence for causal powers when observing causal relations; what he found instead was spatiotemporally ordered successions of events.

So, why do we believe in causal powers? Hume's answer was that knowledge of the causal texture of the world was merely an illusion derived from observed statistical regularities. Illusions can be quite useful--as indeed they have been shown to be for our functioning visual system--but they are a construct of the mind rather than an objective, veridical importation from the physical universe.

Contemporary learning theorists have adopted Hume's empiricist approach in their theories of causal learning. Associations derived from spatiotemporally connected events, such as through Pavlovian and instrumental conditioning, serve in these theories as the basis for causal predictions (e.g., Allan, 1993; Killeen, 1981; Shanks & Dickinson, 1987; Wasserman, 1990). Causal predictions based on covariations between events are deemed sufficient to explain our causal inferences, with no need to resort to the elusive concepts of causal powers or processes.

Hume’s analysis of causality leaves us with a puzzle. His claim seems correct that covariations between observable events are the primary perceptual input for causal inductions. On the other hand, he does not explain why we do not stick to covariations, but try to go beyond the given information by assuming hidden capacities, forces, or mechanisms beyond the surface of orderly event successions (Ahn, Kalish, Medin, & Gelman, 1995; Cheng, 1997). Hume was right when he pointed to covariations as the primary experiential evidence for causal relations. Nevertheless, his empiricist epistemology prevented him from taking the next step. As many philosophers of science have shown, apart from concepts referring to observable events, our theories also contain concepts which are only indirectly tied to the observable data (see Glymour & Stalker, 1980; Quine, 1960; Sneed, 1971). Causal powers may be such theoretical concepts, which people infer based on covariation information (Cheng, 1997).

Why are we unsatisfied with mere covariations? Why are we so interested in causal powers and mechanisms? Different factors may contribute here. Infants may be born with a natural tendency to interpret causal events as caused by hidden forces (e.g., Carey, 2009; Leslie & Keeble, 1987). Other researchers have suggested that the tendency to interpret events causally may be triggered by infants’ experience of their own actions changing events in their environment, which might provide the basis for further causal knowledge (Dickinson & Balleine, 1993; White, 2006). Most likely both factors are at play, but we know very little about their relative contributions.

Regardless of the origin of our tendency to form causal representations, there are a number of computational reasons for the usefulness of causal representations over representations that merely reflect covariations. Recent theoretical developments in philosophy, statistics, and psychology on causal model theory have pinpointed a number of computational advantages of causal models (Cheng, 1997; Glymour, 2001; Gopnik et al., 2004; Lagnado et al., 2007; Pearl, 1988, 2000; Sloman, 2005; Spirtes, Glymour, & Scheines, 1993; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Tenenbaum, Griffiths, & Kemp, 2006; Griffiths & Tenenbaum, 2009; Waldmann, 1996; Waldmann & Holyoak, 1992; Woodward, 2003).

What do those representations of causal models give us that we cannot get from associative knowledge?

(1) **Causal knowledge allows us to accurately represent causal directionality and causal structure.** Although a flagpole perfectly covaries with its shadow, we recognize that the flagpole causes the shadow and not vice versa. By representing this particular causal structure or map, we can make strong inferences about interventions: for example, moving the flag will make the shadow move, but shining our flashlight on the shadow will not affect the flag. By contrast, covariations are undirected and therefore do not allow informed inferences about the outcomes of interventions. A number of experiments with adults have demonstrated that people are sensitive to causal directionality and the causal type (cause vs. effect) of the learning events. Moreover, people are capable of separating representations of causal structure from the temporal order of learning events. That is, causal maps may be inferred in part from temporal maps. Thus, people differentiate between cues which represent causes from cues which represent effects (Waldmann & Holyoak, 1992; Waldmann, 2000, 2001; Young, 1995).

(2) **Causal power.** Covariation is only an indirect indicator of causal power (Cheng, 1997). This point can be most easily seen with ceiling and floor effects. A generative cause will not covary with its effect if the effect is already at its maximal value before the cause is set. Thus, it is possible that an event has causal power, but cannot display it in a measurable covariation. Causal power is a theoretical concept which expresses the strength of a cause in an ideal situation in which alternative causes are absent. In real life, in which we are constantly confronted with complex causal scenarios, we cannot observe such ideal situations. Nevertheless, Cheng (1997) has shown that we can estimate causal power based on covariation, and numerous studies have demonstrated that people often go beyond mere covariations, and try to estimate causal power (Buehner & Cheng, 2005; Buehner, Cheng, & Clifford, 2003; Cheng, 1997; Griffiths & Tenenbaum, 2005; Liljeholm & Cheng, 2007; Waldmann & Hagmayer, 2001; Wu & Cheng, 1999).

(3) **Causal vs. non-causal covariations.** Another important distinction that cannot be made by the covariation view is between causal relations and spurious non-causal relations which both can display equal amounts of covariation. Yellow teeth and lung cancer covary due to a common cause, smoking, although they are not causally related. The importance of this distinction can be best demonstrated when thinking about interventions. Interventions will only work when they target direct or indirect causes of an effect, but they will universally fail when they involve events which only spuriously covary with the outcome. Brushing your teeth with whitening toothpaste will not affect the incidence of lung cancer, whereas quitting smoking will. This distinction allows us to effectively plan actions by choosing only to intervene on causes and to not waste effort intervening on spuriously related events (Woodward, 2003). A number of researchers have shown that adults and children are sensitive to the structural consequences entailed by interventions (Gopnik et al., 2004; Meder, Hagmayer, & Waldmann, 2008, 2009; Sloman & Lagnado, 2004; Waldmann & Hagmayer, 2005).

(4) **Inferring hidden causes.** When a doctor observes nasal congestion, red, watery eyes, swollen lymph nodes, and a cough, he or she can diagnose a viral infection as the probable cause of a common cold. When an infant observes a bean bag being tossed from behind a screen, he or she acts surprised if the screen is subsequently removed to find nobody there (Saxe, Tzelnic, & Carey, 2007). These examples demonstrate the capacity that humans, even very young children, possess in drawing inferences about hidden causes from patterns of observed statistical associations among events. This inference process uses covariations as input, but it allows us to go beyond an associative explanation which can only deal with observable events and their interrelations (Blaisdell, 2008; Gopnik et al., 2004; Kushnir, Gopnik, Lucas, & Schulz, 2010; Waldmann, Hagmayer, & Blaisdell, 2006; but see Blaisdell et al., 2009).

(5) **Causal representations offer the advantage of parsimony.** We would need to encode 15 pairwise covariations in order to learn predictive relations between 6 events. Causal models provide more parsimonious representations. If we know, for instance, that high sugar consumption is the common cause of insulin resistance, wildly fluctuating levels of serum glucose, dental caries (cavities), osteoporosis, and a high body mass index (due to visceral fat storage) (Taubes, 2007), then we can infer all 15 covariations among these events from knowledge about the base rate of the common cause and the five covariations between the common cause and each of its effects (Pearl, 1988, 2000; Spirtes et al., 1993).

(6) **Observation-based and Intervention-based predictions.** One of the most important capacities going beyond associative representations concerns our ability to derive predictions for hypothetical observations and interventions from observational learning data. Associative theories can distinguish between these cases when learning encompassed both observational learning (i.e., classical conditioning) and intervention learning (i.e., instrumental conditioning), but they lack the ability to base these two types of predictions on observational learning input alone. Research on causal Bayes nets (Spirtes et al., 1993; Pearl, 2000; Woodward, 2003) has shown how these predictions can be formally derived (see also Waldmann et al., 2008, for an alternative formalism). Suppose we observe a change in the level of a barometer. We also expect to observe a concomitant change in the weather. This expectation holds because the state of the barometer and the weather are both directly affected by changes in atmospheric pressure (Figure 1, left panel). Thus, all three events covary. If we observe one of the three events, an associative covariation-based theory would predict that we expect the other two events as well. This knowledge could be used to predict events based on observations of other events, but we would be wrong if these other events are caused by an intervention. One fundamental aspect of our causal knowledge is that we know that observing effects allows us to infer the presence of their causes, but manipulating events does not alter their causes, only

> their effects. Thus, if we intervened and tampered with the barometer, thereby artificially altering its reading (Figure 1, right panel), we would not expect a change of the weather. An intervention renders the barometer independent of its normal causes (changes in air pressure) because those causes are no longer setting it – the intervention is setting it (Spirtes et al., 1993; Pearl, 2000). Numerous empirical studies have shown that children and adults can distinguish between observation-based and intervention-based predictions (Gopnik et al., 2004; Meder et al., 2008, 2009; Sloman, 2005; Sloman & Lagnado, 2005; Sloman & Hagmayer, 2006; Waldmann & Hagmayer, 2005; Waldmann et al., 2006; 2008), which can be modeled by causal Bayes nets (Pearl, 1988; 2000; Sloman, 2005; Spirtes et al. 1993; Woodward, 2003).

This short overview demonstrates the computational advantages of causal knowledge over knowledge which merely contains information about covariations. Causal knowledge is not only important when learning about the world, but it also underlies category formation (Lien & Cheng, 2000; Waldmann & Hagmayer, 2006), planning (Pearl, 2000), decision making (Hagmayer & Sloman, 2009; Sloman & Hagmayer, 2006), and moral judgments (Hauser, 2006; Waldmann & Dieterich, 2007).

Causal Reasoning in Rats

One question raised by the extensive evidence that humans engage in rational causal reasoning processes is whether nonhuman animals stick to covariations or can also reason about causation (Gopnik et al., 2004; Gopnik & Schulz, 2007; Kushnir & Gopnik, 2005). Although many developmental psychologists posit that even infants have the capacity for causal representations (see Carey, 2009, for a recent review), many researchers draw a line between human and nonhuman animals, turning causal reasoning into a uniquely human capacity similar to language (Bonawitz et al., 2010; Penn, Holyoak, & Povinelli, 2008; Penn & Povinelli, 2007). For example, Povinelli (2000) claimed that chimpanzees, unlike humans, are incapable of reasoning about hidden forces and causal mechanisms (see also Tomasello & Call, 1997). One shortcoming of this research, however, is that the competencies of chimpanzees were typically tested with relatively complex tasks that require fairly elaborate physical knowledge about mechanisms. It may well be that animals lack such knowledge, but they may still have a basic understanding of the difference between causal and noncausal covariations.

Causal model theory may therefore be a better framework to test whether animals understand the basic features of causality. Causal model theory specifies causal knowledge on a relatively abstract level without requiring deep knowledge about physics. It is certainly possible to have a basic causal understanding of a situation without detailed mechanism knowledge. Basic causal knowledge is present if we can distinguish causes from effects and if we have at least some vague intuitions about causal structure. As shown above, this skeletal knowledge allows us to make numerous interesting inferences. Thus, causal model theory seems better suited to explore causal reasoning in animals than theories of intuitive physics (e.g., Young, Beckmann, & Wasserman, 2006). It may well be that rats, for example, reason causally, but lack knowledge about mechanisms. Moreover, rats almost certainly do not have meta-knowledge about the concept of causality, which can be seen as a sign of advanced scientific reasoning and which may be unique to humans (Penn et al., 2008).

We have recently used predictions derived from causal model theory to explore rat behavior in causal reasoning tasks (Waldmann, Cheng, Hagmayer, & Blaisdell, 2008, for an overview). Causal model theory is a computational account which specifies the goals and capacities of organisms on an abstract level without making claims about the underlying psychological mechanism (cf. Call, 2006; Clayton, Emery, & Dickinson, 2006; Danks, 2008; Kacelnik, 2006; Krechevsky, 1932; Sloman & Fernbach, 2008). It is premature to take any stance regarding the underlying mechanisms, such as propositional (e.g., Mitchell, De Houwer, & Lovibond, 2009) or associative (Castro & Wasserman, 2009; Wasserman, 1990; Young, 1995) representations to name two recently debated possibilities. If the behavior of rats (or any species) turns out to follow the predictions of causal model theory, then this by no means implies that they are consciously or unconsciously using causal Bayes nets in their minds or that the underlying mechanism is symbolic. Our primary question is simply whether rats use representations encoding covariations or whether their behavior reveals that they go beyond covariations toward causal representations. This is an important question even for researchers who are interested in mechanisms. Should rats fall into the second class, then all models of mechanisms are incomplete that cannot take the step beyond covariations.

In this chapter, we review evidence from our laboratory suggesting that rats engage in some forms of causal reasoning that appear to go beyond contemporary associative accounts and that more closely approximate a rational account (see also Beckers, Miller, De Houwer, & Urushihara, 2006; Sawa, 2009). Specifically, we show how rats may reason about the world in a manner consistent with causal model theory. There are also some telling limitations to the rat's ability to approximate a rational causal reasoner which we will also discuss. In the remainder of the chapter, we review research from our laboratory which was conducted to pursue the following objectives: (1) determine if rats can form causal models, (2) establish whether rats understand actions as causal interventions in a rational manner, (3) evaluate what constitutes a good intervention, (4) assess whether rats use interventions to investigate causal structure, and (5) determine whether rats represent hidden events. In future work, we will investigate the cognitive mechanisms underlying causal reasoning in rats, paying particular attention to the role of goal-directed action in reasoning about interventions; we will also examine the neural mechanisms of interventions.

Our research is still in its infancy; thus, the experiments themselves raise many more questions than they answer. They do, however, lay an important groundwork which may serve as a foundation for the comparative analysis of causal cognition. The fruits of this research can provide insight into the fundamental nature of the processes underlying causal cognition, which can help in refining the existing models of causal inference and help in developing new ones. This work also has implications for cognitive science and philosophy by helping to discern the unique elements of human psychological processes from those that are shared with other species.

Our experiments involve Pavlovian and operant conditioning procedures administered in commercially available rodent test chambers (Figure 2, Med Associates, Georgia, VT). These chambers are equipped with three speakers which can be used to present auditory stimuli (e.g., tones, white noise, click trains), an incandescent house light and a diffuse light for the production of visual stimuli, two retractable levers (right and left) which can be inserted into or withdrawn from the chamber, and a food niche where sucrose solution (20%) can be delivered. Our subjects are female Long-Evans rats purchased from a commercial vendor and maintained at 85% of their free-feeding weight to provide motivation for food-seeking behavior in the test chamber. Independent manipulations include presentations of audiovisual stimuli and pairings of these stimuli with sucrose solution. Dependent measures include lever-pressing behavior and nose poking into the food niche. Nose poking serves as a proxy measure of expectation of the delivery of food (sucrose solution). Thus, predictions of high and low expectations of food lead us to predict high or low rates of nose poking, respectively. Although not discussed explicitly below, independent factors were counterbalanced appropriately (e.g., which of two audio cues served in a particular functional role or which of two levers served a particular function).

Do Rats Form Causal Models?

Our first experiment deployed the basic procedure that we have used throughout our research to teach causal models to rats (Blaisdell, Sawa, Leising, & Waldmann, 2006). Once established, this procedure allowed us to assess whether rats use these causal models to reason rationally (see next section).

We used Pavlovian pairings of a Light with Tone (Light→Tone) and Light with Food (Light→Food) to teach the rats a common cause model, in which a Light was a common cause of both Tone and Food (left panel of Figure 3 – analogous to how air pressure is a common cause to both changes in the barometer and changes in the weather, Figure 1). We did so by presenting two types of trials within each training session. On one type of trial, a diffuse Light was flashed on and off for 10 seconds and after it terminated a steady Tone was presented for 10 s. The second type of trial consisted of similar presentations of the flashing Light followed upon its termination by 10-s presentation of Food (sucrose solution) by raising the dipper containing sucrose into the food niche for 10 s after which it was removed from the niche. We expected rats to form a common cause model representation through these learning trials (additional procedural details can be found in Blaisdell et al., 2006).

Our learning procedure raises the question whether it is the appropriate technique to teach rats a common cause model. Whereas common cause models imply that the effects of the common cause (e.g., Tone, Food) are positively correlated, our sensory conditioning procedure presented these events as negatively correlated; Tone and Food always occur in the absence of each other. It is, however, well established that learning trials such as the ones we have chosen may lead—at least with a small number of such trials—to sensory preconditioning, which is the excitatory response to an initially behaviorally neutral stimulus that had been paired with another initially behaviorally neutral stimulus that subsequently was conditioned as a CS (Brogden, 1939; Pavlov, 1927; Yin, Barnett, & Miller, 1994). To prevent the rats from directly experiencing a positive correlation between Tone and Food, which might cause them to induce a *direct* causal relation between Tone and Food, we chose to use the sensory preconditioning procedure which *indirectly* links Tone and Food through its common cause, the Light. Thus, we predicted that rats will learn about the direct causal relations between Light and Tone and Light and Food during the learning trials, and infer a positive correlation between the indirectly linked events Tone and Food without paying attention to the negative correlation in the learning input. This representation would then be consistent a common cause model (see also Waldmann et al., 2008). Note that this hypothesis can be empirically tested by looking at whether Tones elicit excitatory or inhibitory expectations of food. Our experiments clearly provide evidence that rats learn to expect food after the Tone following second-order conditioning training.

All rats also received a third type of trial within each training session, consisting of simultaneous pairings of a 10-s Click train with 10-s delivery of Food (sucrose). We used simultaneous pairings because prior work has found rats to be highly sensitive to the temporal relationship between events in sensory preconditioning (Leising, Sawa, & Blaisdell, 2007; Savastano & Miller, 1998). For example, in an appetitive sensory preconditioning experiment with rats in which a 60-s tone was paired with a 10-s light in Phase 1, and the 10-s light was simultaneously paired with food in Phase 2, rats were observed nose poking the most during a subsequent test of the tone at the time that the food would be expected if rats had integrated the tone-light and light-food temporal intervals (Leising et al.). If rats in the current experiment integrated the light-tone and light-food temporal intervals during training, then rats should expect food during the tone at test. Furthermore, because we wished to equate the time at which rats in the current experiment expected food during test trials with the Tone and the Click, we decided to present the click and food simultaneously during training. The simultaneous Click-Food trials established Click as a direct cause of Food and served as a control manipulation (see below).

Evidence that rats learned the second-order Tone-Food relationship came from a subsequent test phase in which rats were presented with the Tone, but without presentations of Light or Food (Observation test). [Note, each rat in the Observation test condition was yoked to a Master rat in an Intervention test condition described below. Each time the master rat pressed a lever in its chamber, both the master rat and the yoked rat received a presentation of the tone. We used this yoking procedure to equate the number and timing of tone presentations at test between the test conditions.] The results showed that the Tone prompted the rats to expect food delivery, which was measured by the high rate of nose poking into the food niche (Figure 4, black bar for Condition Common Cause). This behavior is consistent with the view that the rats accessed a common cause model to infer from one effect (Tone), through the Light, to the other (Food). This account is similar to meditational learning, in which an event such as the light can mediate conditioning to another event with which it is associated (e.g., Tone) (Holland, 1990). The amount of responding during observation tests of the Tone, which had a second-order relationship to the Food during training, was equivalent to the level of responding during observation tests of the Click, which had a first-order simultaneous relationship with the Food during training (Figure 4, black bar for Condition Direct Cause), suggesting that both test stimuli elicited similar levels of expectation of food. [Each rat in the Observe test condition received presentations of the click at test whenever a rat in the Intervene test condition pressed a lever in its chamber.]

Do Rats "Do"?

Nose poking during the Tone in the Observation condition can be interpreted as evidence for the formation of a common cause model, but is also consistent with the hypothesis that rats had formed a second-order associative relationship between the Tone and the Food (Pavlov, 1927; Yin, Barnet, & Miller, 1994). Thus, the crucial test of causal model theory requires a way to distinguish between predictions made by associative accounts from those made by causal model theory.

One of the crucial distinctions discussed in the Introduction was that causal model theory predicts that subjects should be sensitive to whether the event was merely observed ("seeing") or was produced by an intervention ("doing"). According to this theory, the passive observation of an event should be represented differently from the same event being caused by an intervention (Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). An intervention onto an event should lead to the inference that this event is separated from its previously established causes—what Pearl (2000) terms graph surgery and what cognitive psychologists call discounting. It has previously been shown that nonhuman animals are capable of discriminating between outcomes which were caused by their own actions and outcomes which were not caused by their own actions (e.g., Killeen, 1978; 1981)—a necessary prerequisite ability for inferences drawn from interventions.

Our central question was whether rats' causal inferences are sensitive to the distinction between events that were merely observed and events that they have caused by means of an intervention. To test this competency, we (Blaisdell et al., 2006) introduced a novel lever in the test phase (Figure 3, right panel). It is important to note that this is the first time the rats had seen this lever. They had never seen it during training and they had not had any experience pressing the lever prior to the test phase. (Note that rats in both the Intervene and Observe test conditions had a lever, but for rats in the Observe test conditions, the lever was non-functional: that is, pressing the lever had no stimulus consequences of any kind.) In the Intervene test condition, the 10-s Tone was presented whenever the rat happened to press the lever. We recorded nose poking into the food niche (measured by breaks in a photo beam projected across the entrance to the food niche) during each 10-s presentation of the Tone during the test session as our assessment of expectation of food delivery.

According to causal model theory, if the rats understood that their intervention (and not the Light) produced the Tone on the Intervention test, then the Tone should not lead them to expect Food to be available (Figure 5, top-left panel). This possibility is analogous to our reasoning that, when we tamper with the barometer, we should not expect the weather to change (Figure 1, right panel). Indeed, the rats that turned on the Tone through their intervention on the lever, nose poked during the Tone much less than did the rats receiving the Observe test (Figure 4, gray bar for Condition Common Cause, modeled in Figure 5, top-right panel). Note that due to the yoking procedure, the rats in the Observe test condition received an equal number of test trials with the Tone as did subjects in the Intervene test condition, equating stimulus exposure in the two testing procedures. The statistical relationship between Tone and Food that was experienced during training was identical in the Intervene and Observe test conditions; thus, an associative account would predict equal amounts of nose poking in both test conditions. Our results, therefore, are consistent with causal model theory, but not with a conventional associative account.

To show that lever pressing behavior did not simply interfere with nose poking, two more groups of rats were tested on the Click instead of (and in a similar manner as) the Tone (Figure 3, right panel). Causal model theory predicts that direct causes should lead to the expectation of their effect regardless of whether they were generated by an intervention or merely observed (Figure 5, right panels). Interestingly, in this test situation, nose poking during the Click that was produced by a lever press (Intervene test) was not lower than nose poking during a Click that was merely observed (Observe test; Figure 4, right-hand bars). If pressing the lever interfered with nose-poke responding (i.e., if the rat could not be doing both actions at the same time), then we should expect a similar disruption of nose poking during the Click by the lever press. Thus, the pattern of results of Experiment 1 of Blaisdell et al. (2006) is fully consistent with the predictions of causal model theory, but not with those of associative accounts.

If lever pressing attenuates nose poking through response competition, then lever pressing and nose poking during each trial should be negatively correlated (that is, the more lever presses recorded on a trial, the fewer nose pokes should be observed). Analysis of the correlations between mean trial lever presses and mean trial nose pokes in the Intervene conditions of Experiment 1 of Blaisdell et al. (2006) fails to support a response-competition account of the effect of the lever-press intervention in the

Common Cause condition (r^2 = 0.085, p > .38, Figure 6 top panel, diamond symbols). Correlations were also not negative in the Direct Cause condition (r^2 = 0, p > .98, Figure 6 top panel, square symbols). Thus, there is no evidence that the difference in how lever pressing affected nose poking in the Common Cause versus Direct Cause conditions was driven by response competition.

Further Evidence against Response Competition

Dwyer, Starns, and Honey (2009) replicated the procedures of Blaisdell et al. (2006). They found, however, that nose poking was typically lower during trials in the Intervene test conditions than in Observe test conditions. This was the case both for the Tone (Common Cause condition) and Click (Direct Cause condition) in their Experiment 1. We cannot tell at this point why Dwyer et al. failed to replicate the interaction between observing/intervening and the causal model which is crucial for our conclusion that rats reason causally, and which we have found in several experiments (see also below). Nevertheless, Dwyer et al.'s alternative interpretation may not explain our results nor their own.

Dwyer et al. (2009) argued that their results are best explained by response competition. They also suggested that such an account may plausibly apply to the results of Blaisdell et al. (2006). We have already shown above that there was no evidence for response competition in our experiment. Moreover, a response competition account is inconsistent with the interaction that we obtained in our studies. In fact, a closer inspection of the results from Experiment 1 of Dwyer et al.'s study reveals some inconsistencies with the response-competition interpretation. Figure 7 shows the results of the replication by Dwyer et al. (2009; Experiment 1) of the 2 days of testing as in

Blaisdell et al. (2006). They reported a small, nonsignificant difference in nose poking during the Tone between the Intervene and Observe test conditions for subjects tested on the Common Cause model, but significantly less nose-poke responding in the Intervene than Observe test condition for subjects tested on the Direct Cause condition. This pattern of results is exactly opposite from that reported by Blaisdell et al. (see Figure 4). Dwyer et al. attribute their results to response competition between lever pressing and nose poking during test trials. An analysis of the correlations between mean test trial lever presses and mean test trial nose pokes (raw data supplied by D. Dwyer, personal communication), however, fails to support this interpretation. Correlations were not significantly negative in either the Common Cause test condition ($r^2 = 0.029$, $p > .50$, Figure 6 bottom panel, diamond symbols) or in the Direct Cause test condition ($r^2 = .008$, $p > .73$, Figure 6 bottom panel, square symbols). Although we have yet to determine the source of the puzzling difference in patterns of results obtained by Dwyer et al. and Blaisdell et al., in neither case do we believe that response competition can serve as a plausible account.

Another possible strategy to rescue an associative account of the findings in Experiment 1 might be to argue for differences in Tone-Food and Click-Food associative strengths. Click had a first-order relationship to the Food, whereas Tone had a second-order relationship to the Food. Second-order Pavlovian events are often noticeably weaker than first-order events (but see Barnet, Cole, & Miller, 1997; Barnet, Grahame, & Miller, 1991; & Cole, Barnet, & Miller, 1995 for exceptions). If the lever press intervention did exert some interference, then it would more likely affect the second-order Tone than the first-order Click.

Although the amount of nose poking during test trials of the Tone and Click in Experiment 1 (Observe test conditions) were roughly equivalent, it may still be the case that the underlying Click-Food association was stronger than the Tone-Food association and therefore less subject to interference from lever pressing. Thus, in Experiment 2, we compared two sensory preconditioning preparations which, like Experiment 1, predicted different effects for the Intervene and Observe test conditions. We compared the Common Cause condition (as in Experiment 1) with a Causal Chain condition (Figure 8). Causal chain training consisted of trials on which Tone was forward paired with Light (Tone→Light) in Phase 1 of sensory preconditioning, and then Light was forward paired with Food (Light→Food) in Phase 2 of sensory preconditioning. Common cause training was similar to Experiment 1 except that Light-Tone trials were given all in Phase 1 and Light-Food trials were given all in Phase 2. Causal model theory predicts the same results for causal chains as for direct causes. If an indirect cause of an effect is produced by an intervention or observed, both the intermediate cause (Light) and the final effect (Food) should be expected.

At test, the novel lever was inserted into the chamber and lever presses turned on the Tone in the Intervene test conditions, but had no consequence in the Observe test conditions. If lever pressing disrupts nose poke responding to events merely because they have a second-order relationship to Food (and thus are weaker than a first-order stimulus), then nose poke responding during the Tone at test should be disrupted by the lever press in both the Common Cause Intervene and Causal Chain Intervene test conditions. If, however, lever pressing disrupts nose poking during the Tone through discounting, then nose poking during the Tone should be disrupted in the

Common Cause Intervene test, but not in the Causal Chain Intervene test (compare left and right panels of Figure 8). Blaisdell et al. (2006, Exp. 2a) found that lever pressing did not disrupt nose poking during the Tone in the Intervene test condition for rats that received causal chain training (Figure 9, right panel). They replicated a strong attenuating effect of the lever press intervention on nose poking during the Tone in the rats that received Common Cause training, as seen in Experiment 1. This difference in how the lever press affected nose poking during the Tone in the Intervention conditions therefore fails to support the view that the lever press intervention attenuated responding to the Tone in Experiment 1 merely because it had a second-order relationship to Food. Rather, these results are consistent with the view that the rats acted as if they understood the causal relationship between their action and an outcome (cf. Killeen, 1978). If in the Chain conditions, the Tone was represented as a cause of the Light, which was represented as a cause of Food, then it should not have mattered whether the Tone was merely observed or caused by an intervention; the Food should have been expected in either case. Although nose-poke responding in the Observe test condition was lower in the Chain condition than in the Common Cause condition, responding was nevertheless significantly higher than for a third set of rats for which the Light had not been paired with Food during training (Unpaired conditions in Figure 9)--(see, however, Dwyer et al., 2009, for diverging findings on a chain structure; see discussion above).

What's so Special about Actions?

Our experiments provide supportive evidence for causal inferences in rats that associative theories fail to explain. One key claim of causal model theory is that the default assumption about many of our interventions is that they are *independent* of other events and that they deterministically fix the states of the variable that our intervention directly targets (e.g., the state of the lever)(Waldmann, Hagmayer, & Blaisdell, 2006).

Deterministic causes are more readily perceived as being causal than are probabilistic causes. The cause of a plant’s death is more readily apparent when it is yanked out by the roots and left on the ground than if it has been soaked by a strong rainstorm. Getting a flu shot may or may not be an effective prophylactic against catching the flu (current evidence is not overwhelming), but avoiding contact with any person or surface infected with the flu virus is a guaranteed prevention.

A cause that produces an effect in the absence of other confounding causes is easily recognized as having an independent unconfounded causal influence on its effect. Causal relations can be readily induced on the basis of interventions that act independently of the causal system on which they act (Woodward, 2003). Unlike externally observed events, self-generated actions typically are seen as independent (Killeen, 1978, 1980), which would allow the actor to infer causality after very few (or even one) learning trials. This is the reason why experimental outcomes are always preferable to epidemiological correlations in establishing cause-effect relationships in science and medicine.

Self-generated actions are often viewed by agents (the actor) as both deterministic and independent; they may thus hold a special status for deriving causal knowledge. “If I flick this switch, the light turns on. If I don’t flick the switch, the room remains dark.” Such a simple cause-effect relationship can readily be determined through intervention, whereas observations typically necessitate consideration of possible confounding factors. (Note, the agent does not have to actively intervene; the agent can merely observe another agent intervening or observe a fortuitous intervention, such as a book falling off the shelf and accidentally flicking the switch on its fall to the floor. See discussion by Tomasello & Call, 1997.)

The ability to reason about cause-effect relationships through an intervention on a single variable is the basis for the scientific method, which gives humankind an incredible analytic power. “If I put reagent X into a beaker filled with reagent Y, the mixture ignites, otherwise the mixture remains inert.” “If I look at a blue-filled circle for a sufficient duration, I then see a yellow, circular after-image when I look at a white wall, but not otherwise.”

We recently found evidence that rats treat their actions as special (Leising, Wong, Waldmann, & Blaisdell, 2008). We compared the efficacy of an action to that of a salient exogenous cue under conditions in which both were equated in their contiguity and contingency with one of the effects (a Tone) of a common cause model. All rats first received common cause training (Phases 1 and 2) as in Blaisdell et al. (2006, Exp. 2a). Then rats were allocated to one of three test conditions (see design in Figure 10 bottom panel). Rats receiving the Intervene test were presented with a 10-s Tone every time they pressed the lever (except for lever presses that occurred while the Tone was already on, which had no consequence). Each rat receiving the Observe test was yoked to a rat in the Intervene test condition, so that the Observe rat received a Tone every time the Intervene rat did. Thus, these two conditions replicated the Intervene and Observe test conditions of Blaisdell et al. In the third test condition (Exogenous Cue) each rat was also yoked to a rat in the Intervene test condition. Every time a rat in the Intervene test condition received a Tone (because it pressed the lever), the rat in the Exogenous Cue condition received a presentation of a novel stimulus (a Click) followed by the Tone. In Experiment 1 of Leising et al. (2008), the Click remained on for 10 s and was followed on its termination by the 10-s Tone. In Experiment 2 of Leising et al., the Click started as soon as the Intervene rat to which it was yoked pressed the lever and terminated as soon as the Intervene rat stopped pressing the lever. Upon the termination of the Click, the 10-s Tone was then presented. Thus, in Experiment 1 the duration of the Click matched the duration of the 10-s Light that had been paired with the Tone during training, and in Experiment 2 the duration of the Click matched the duration of the lever press by the Intervene rat on each test trial. The question was whether the Click would be as effective as the lever press in leading rats to discount the causal influence of the common cause Light (compare top-left and top-right panels of Figure 10).

Figure 11 shows the results of both of these experiments. Consistent with the predictions of causal model theory, we found that the lever press, but not the exogenous cue led to discounting of the common cause Light. This finding is consistent with the assumption that actions have a privileged role as interventions for rats.

One further important attribute of interventions concerns possible transfer effects. Causal model theory predicts that inferences drawn from an intervention should be restricted to the moment of action and should not transfer to later tests in which the intervention is absent. Thus, a causal reasoner should be capable of switching back and forth between inferences based on actions or observations without being influenced by previous predictions. Unlike most associative processes, therefore, reasoning from the presence or absence of interventions should be path independent.

For example, if I water my front yard and then notice that the sidewalk is wet, I infer that it was I and not rain that caused the wet sidewalk. If on the very next day, however, I notice that the sidewalk is wet and I know that I haven't watered my lawn that day, then I infer that it must have recently rained. Discounting of rain occurs in the instance in which I intervened, but that inference does not carry over to the next day on which I did not intervene; thus, no discounting is expected (see related example by Clayton & Dickinson, 2006).

Leising et al. (2008) tested whether rats understand this principle of interventional reasoning with a study replicating the training conditions from Blaisdell et al. (2006, Exp. 1) in which all rats were trained on both the common cause model (Tone←Light→Food) and direct cause model (Click-Food), except that the Light→Tone trials were all given in Phase 1, and the Light→Food and Click-Food trials were all given in Phase 2 (see design in Figure 12). Half the rats then received testing on the Tone (Common Cause test conditions), while the remainder received testing on the Click (Direct Cause test conditions). Each rat received one test session of the Intervene test condition and a second test session with the Observe test condition (test order counterbalanced).

Figure 13 shows that rats can flexibly switch between responding inferentially to a Tone that was observed versus a Tone that was the result of their intervention. Hence, a rat receiving training on the common cause model (Tone←Light→Food) that intervened on the Tone on Day 1 of testing – which reduced its expectation of Food – had an increased expectation of Food on Day 2 of testing when they merely observed the Tone. Importantly, this finding means that exposure to the contingent relationship between a lever press and the Tone, which could lead to the acquisition of a Lever Press→Tone association in the first test session, did not transfer to affect responding on the second test session. Likewise, a rat that had merely observed a Tone on test Day 1 – thus leading it to expect Food – had a lower expectation of Food on test Day 2 when it intervened to produce the Tone. As expected, a lever press intervention did not affect food-related responding (nose poking) when rats were tested on a Click that had been established as a direct cause of the Food, thereby replicating the results of Experiment 1 of Blaisdell et al. (2006) and contradicting the results of Dwyer et al. (2009).

One last piece of evidence supporting causal reasoning in rats follows from the immediacy of the causal inference derived from an intervention. We have a profound sense of causality when we accidentaly bump into a table and thereby spill a glass of wine, probably because we regard our own actions as unconfounded (Woodward, 2002). We do not need multiple observations of this relationship to realize that we caused the wine to spill; it is immediately apparent on the very first instance. Do rats likewise reason similarly about their impromptu effects on the world?

To answer this question, we performed a metaanalysis of first-trial test performance of nose pokes across all of our reported data sets involving common cause training and Intervene and Observe testing to determine if the effect of an intervention on the expectation of Food is present on the very first test trial. Appetitive conditioning in which an external stimulus, such as an audio or visual cue, signals delivery of Food in a food niche under conventional parameters typically takes dozens of trials before food-seeking behavior comes under control of the cue (Gallistel & Gibbon, 2000). Moreover, inhibitory behavioral control by a conditioned inhibitor takes an order of magnitude longer than by a conditioned excitor to develop any measurable behavioral control (Yin, Barnet, & Miller, 1994). Moreover, according to most contemporary models of associative learning, the effect of the first learning trial should not be apparent until the second presentation of the Pavlovian stimulus or instrumental response (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). Figure 14 presents the metaanalysis by Leising et al. (2008) which shows a strong attenuating effect of a lever press intervention on nose-poke responding during the Tone on the first trial on which the rat intervened to produce the Tone. Thus, rats seem to understand on the very first trial that their novel actions are causal. These results are rational under the assumption of causal model theory that actions are typically viewed as independent and deterministic, but pose quite a challenge for associative models.

From Reasoning to Acting?

Children and adults use many cues to infer causal structure, such as temporal and spatial contiguity (Leslie & Keeble, 1987; Michotte, 1946/1963), temporal priority (Hagmayer & Waldmann, 2002), covariation and contingency (Cheng, 1997), and prior experience (see reviews by Lagnado et al., 2007; Young, 1995). But, perhaps the most powerful and effective guide to causal structure comes from data produced by interventions – an idea that is currently transforming the fields of statistics, philosophy, computer science, and psychology (Gopnik et al., 2004; Woodward, 2002). Indeed, interventions are the primary means by which scientists can differentiate causal relationships through experimentation from mere observed patterns of correlations. The core idea is this: knowing that X directly causes Y means that, all else being equal, intervening to change X can change Y (Pearl, 1988, 2000; Spirtes et al., 1993; Woodward, 2003). Adults (Waldmann & Hagmayer, 2005), children (Gopnik et al., 2004; Schulz, Gopnik, & Glymour, 2007), and, as we have shown above, even rats (Blaisdell et al., 2006, Leising et al., 2008) are able to make correct causal predictions about interventions.

So far, we have focused on studies in which rats made observational and interventional predictions after having received purely observational learning input about causal models. Allowing an organism to actually intervene during the learning phase may also provide aid for inducing the correct causal model. To take an example from the introduction, assume, for example, you come from Mars and have no prior knowledge about the causal model underlying smoking, lung cancer, and yellowed fingers (adapted from Gopnik et al., 2004). What should you assume about the causal relationship among this set of three variables if only covariation, but no temporal information is available? In Figure 15, a number of alternative models are depicted that are equally consistent with the learning input. One efficient strategy to constrain the set of models is to use interventions. You could, for instance, continue to smoke but wash your hands or you could color your hands yellow and not smoke. If lung cancer occurred in the former case but not the latter, then this result would strongly indicate that smoking and not yellow hands was the direct cause of lung cancer. Reasoning about interventions is the basis of the scientific method and of much everyday learning.

The ability to reason about interventions has an important functional implication: knowledge of cause-effect relationships should enable one to use interventions on the cause to bring about a desired effect. Schulz et al. (2007) showed that even young children (4-5 years old) – through play with a simple toy box with a switch and two gears – could discover the correct causal model representing the toy's mechanism.

In our experiments, rats merely passively observed the causal model. The interventions in the test phase were just used as a tool to investigate rats' inferences. An interesting question is whether rats would transfer the knowledge that they acquired in an observational context to their action system. If both systems were integrated, then rats should start pressing the lever more often if they were hungry and they thought that lever pressing causes food. They should, however, refrain from pressing the lever if there was no causal, but just a spurious non-causal relation to the food. Blaisdell et al.'s (2006) experiments provided an opportunity to investigate this question. Based on the assumption that observational and interventional learning are integrated, we might expect that rats that intervened on the Click in the Direct Cause condition of Experiment 1, or that intervened on the Tone in the Causal Chain conditions of Experiments 2a and 2b, should press the lever more often than should rats that intervened on the Tone in the Common Cause condition. Moreover, we would expect rats in the Direct Cause and Causal Chain, but not in the Common Cause conditions receiving the Intervene test, to press the lever (which resulted in the onset of the Tone) more than rats receiving the Observe test condition for which pressing the lever did not result in the onset of the Tone. Surprisingly, the experimental results indicated that subjects in the Common Cause condition and Direct Cause condition of Experiment 1 made roughly the same number of lever presses (Figure 16). Thus, rats failed to transfer observationally-learned causal knowledge to their action system. This finding is especially remarkable given the consistent finding that rats made differential predictions of Food based on interventions on an effect versus on a cause. If this failure holds up under further scrutiny, then it highlights a very interesting, and important constraint on the cognitive processes underlying causal cognition in the rat.

Whether human infants can freely transfer observational knowledge to actions is not yet known. It is interesting, however, that the ability to make inferences about hidden objects develops earlier in human infants than does the ability to reach for the hidden object—although infants are already capable of reaching for visible objects (Munakata, 2001; Munakata, McClelland, Johnson, & Siegler, 1997; Munakata & Yerys, 2001). Also, toddlers appear capable of using predictive relations between physically connected events to initiate causally effective actions only when the relationship between the connected events was described to them in causal language (Bonawitz et al., 2010). With age, children eventually develop the capacity to base action selection on causal representations, as do adult humans. Rats have so far not shown this ability. Thus, observational and interventional learning may represent separate systems in both rats and infants, which, at least in humans, are later integrated into unified causal representations.

Hidden Event Cognition

We rarely have direct access to all of the information about causal relationships that govern any particular system. A doctor can merely observe a patient present with red and watery eyes, a runny nose, swollen and red tonsils, and a low-grade fever to infer a hidden viral cause of these symptoms. Likewise, it was the odd, unpredicted movements of Uranus that led Alexis Bouvard in the early 19th century and later Urbain Le Verrier in 1845—both using the physical-causal system of Newtonian Mechanics—to postulate the existence of the as yet undiscovered planet Neptune.

If rats form causal representations, as the evidence seems to suggest, then to what extent do rats draw inferences about hidden causes? It turns out that when we first conducted the chain experiment described above (Blaisdell et al., 2006; Experiments 2a and 2b), we made a startling discovery. During sensory preconditioning training of the causal chain, rats received pairings of the Tone followed by the Light (Tone→Light) and then pairings of the Light followed by Food (Light→Food). As we showed, this learning established a Tone→Light→Food causal chain in which the Tone is a cause of the Light which in turn is a cause of the Food. We were surprised to find very little nose poking in an unpublished experiment in which rats were tested with the Tone (Observation test condition).

At first, we thought that causal chain training had failed. In a subsequent test, however, we removed the light bulb with which the Light had been presented during training. We did so based on the notion that perhaps upon hearing the Tone at test, the rats expected to observe the Light illuminate. The Light did not illuminate at test, however, thereby violating the rats’ expectation. This failure for the light to illuminate in turn may have violated their expectancy that Food would be delivered (because it had always followed delivery of the Light during training), and therefore the rats did not nose poke in the feeding niche. By removing the Light, it became ambiguous as to whether or not the Light was on following the Tone. Because the Light had always followed the Tone during Phase 1 of sensory preconditioning training, the Tone→Light contingency would plausibly lead the rats to expect the Light to be present following the Tone at test, despite the fact that they could not verify the status of the Light. The rats, therefore, might also continue to expect the Food to be present during testing with the Tone, in which case they should nose poke.

It turned out that, when we directly tested this prediction (Blaisdell, Leising, Stahlman, & Waldmann, 2009), the rats did nose poke when the Light was absent, but not when the (unlit) Light was present during Observation tests with the Tone (Figure 17, top panel). In fact, nose poking in the Light-absent condition was significantly greater than in the Light-present condition (black bars), and also significantly greater than an unpaired control group (gray bars) for which the Light had not been paired with Food during training (and thus, the Light should not have raised the expectation of Food).

This surprising result was our first indication that rats distinguish between the explicit absence of an event and uncertainty about the invisible event's status due to lack of information. That is, like human adults (Hagmayer & Waldmann, 2007) and even children (Kushnir et al., 2010), rats are sensitive to the conditions under which they should be able to observe an event and those conditions under which the event should be hidden from observation.

It could be argued, however, that the removal of the light bulb at test created a different context from that in which sensory preconditioning treatment had been administered during training. Note that, as sensory preconditioning actually entails a negative contingency between the second-order Tone and the Food, the second-order Tone may accrue both excitatory and inhibitory properties (suggestion offered by Tom Beckers, personal communication). To the extent that removal of the light bulb at test introduces a context shift, it is possible that the excitatory properties that accrued to the Tone during training transferred to this new context more readily than did the inhibitory properties that may have accrued to the Tone (i.e., a renewal effect, Bouton, 1993; but see Bouton, 1994 for evidence against greater context sensitivity of inhibition than excitation to ambiguous stimuli).

To discriminate between these alternative accounts, we replicated the design of Blaisdell et al. (2009) with the additional manipulation of testing the Tone in the same context as training or in a different context that was explicitly made to be dramatically different from that experienced during training. Rats were allocated to four groups: Present-Same, Absent-Same, Present-Different, and Absent-Different. Present versus Absent indicates whether the Light’s bulb was present or absent during the test on the Tone. Same versus Different indicates whether the test context was the same as or different from the training context. If the reason Blaisdell et al. (2009) found higher rates of nose poking in the Light Absent test condition was due to a context shift created by removal of the light bulb, then by explicitly rendering the test context dramatically different than the training context, we should observe high rates of nose poking in both groups tested in the different context, irrespective of the presence or absence of the light bulb.

Figure 17 (bottom panel) reveals that testing in a very different context actually resulted in relatively little nose poking compared to pre-Tone baseline rates of nose poking. Only in the condition in which the Tone was tested in the same context as used for training, but with the light bulb removed at test did we observe nose poking to be significantly above baseline rates. These results replicate the initial findings of Blaisdell et al. (2009) and refute the context-shift (i.e., renewal) account of their findings.

Recently, we also started to explore a simpler paradigm designed to show that rats distinguish between the explicit versus ambiguous absence of anticipated events (Waldmann, Schmid, Wong, & Blaisdell, 2010). We used an extinction paradigm in which a light was first consistently paired with food and then was extinguished by being presented in the absence of food. The crucial manipulation involved information about the absent food in the extinction phase. Whereas in the Cover condition informational access to the food niche was covered by a metal plate, in the No Cover condition, which represents standard extinction manipulation, the food niche was accessible. Notably in both the Cover and No Cover condition, light was followed by the absence of food. The only difference was whether the food was explicitly (No Cover) or ambiguously (Cover) absent. The test phase (in which the food niche was uncovered for all animals) revealed higher rates of nosepoking in the Cover than in the No Cover condition, suggesting that rats in the Cover condition had higher expectations of food than did rats in the No Cover condition. This difference is consistent with the hypothesis that the rats were able to understand that the cover blocked access to the outcome information, and therefore the changed learning input did not necessarily signify a change of the underlying contingency in the world.

An alternative explanation is that the greater amount of nose poking observed in Group Cover was due to the renewal of excitatory responding after the cover was removed. That is, the introduction of the novel cover during extinction could have created a different context. It is well established that extinction in one context does not generalize to other contexts as readily as does excitatory conditioning (Bouton 1993). We tested this alternative account in a follow up experiment that included an additional Cover Control group of rats that also had a novel cover introduced only during extinction treatment. For this group, however, instead of being placed over the food niche, the cover was placed next to the food niche. Group Cover Control therefore had the same nominal contextual change during extinction treatment as did Group Cover, but without obstructing the food niche. It turned out that Group Cover Control nosepoked as little during the CS at test as did Group No Cover, and significantly less than did rats in Group Cover. Thus, the effect of the cover on protection from extinction cannot be accounted for by the renewal effect.

Conclusions

Rats are capable of using covariation information to form causal representations. These representations include direct cause-effect relationships as well as higher-order causal maps acquired through higher-order associative procedures, such as sensory preconditioning and second-order conditioning (Blaisdell, 2009). The findings in the experiments by Blaisdell et al. (2006) and Leising et al. (2008) reviewed in this chapter support the framework of causal model theory as an account of rats’ learning, whereas they challenge the computational power of contemporary models of associative learning. The most important evidence for this claim is that rats distinguish in their predictions between states of predictive variables that were merely observed versus caused by an intervention. This distinction was shown when rats discounted (i.e., did not act on) a second-order relationship between a cue (Tone) and Food if the cue was the product of their intervention by a lever press and not merely observed in the absence of their intervention. Importantly, we found discounting only when the cue was an effect of a common cause, but not when it was either a direct cause of food (Blaisdell et al., 2006, Experiment 1) or when the cue was a second-order cause in a causal chain (Blaisdell et al., 2006, Experiment 2). This pattern of results supports causal model theory, but it is inconsistent with associative accounts. Furthermore, the analysis of the correlations between lever pressing and nose poking during test trials failed to find any evidence in favor of a simple response competition account.

Interventions were shown to be most effective when they consisted of an action, such as pressing a lever, compared to exogenous events, such as a novel auditory stimulus (Leising et al., 2008, Experiments 1 and 2). Furthermore, discounting was observed only at the moment during which an individual rat intervened with a lever press (Leising et al., 2008, Experiment 3). No carry-over effects of exposure to the lever-press→Tone relationship were observed on subsequent tests of the Tone alone (in the absence of a lever press intervention). This flexibility in turning on and off the discounting effect is another hallmark of causal reasoning.

Finally, a meta-analysis of first-trial performance revealed discounting by a lever press intervention on the first encounter with the novel contingency between the lever press and the Tone (Leising et al., 2008). This result provides further support for causal model theory.

Most of our studies have focused on showing that causal reasoning in rats is consistent with the unique predictions of causal model theory. We also have discussed evidence, however, that demonstrates limitations in rats' powers of reasoning. Although the evidence is still preliminary, and further research is planned, rats seem to have difficulty transferring their observationally gained knowledge to their action system. Although they correctly, in the framework of causal model theory, differentiated between observing and intervening in their expectation of food depending on whether the test cue was part of the common-cause map on the one hand, or the direct cause or causal chain map on the other, they did not adapt their actions to this knowledge. It may be that the procedures used in the two reported experiments performed in our laboratory were not sensitive enough to uncover this ability or it may be that rats truly lack this ability altogether. We plan further research to address this issue. Nevertheless, if causal knowledge in rats is tied to the system used to acquire it (e.g., observations or actions), then interesting questions are raised about the quality of rats' causal reasoning and the underlying psychological and neural mechanisms (cf. Bonawitz et al., 2010 for a similar analysis applied to causal inferences in young children).

In the final empirical section of this chapter, we reviewed evidence from our laboratory that rats distinguish between the explicit absence of a visual event and uncertainty about the state of an unobserved event due to lack of information (Blaisdell et al., 2009). Rats were exposed to causal chain training resulting in the formation of the causal chain Tone→Light→Food. When presented with the Tone at test, rats expected Food (assessed through nose poke responding) more if the bulb on which the Light had been presented during training had been removed from the test chamber during testing than if the bulb remained in the chamber (although unlit). The rats acted as if they realized that the visible light should be on after the tone, whereas they seemed to understand that it may be on, just not be visible, when the light has been removed from sight. We also studied this competency in a simpler paradigm. In an extinction study, information about the light's absence was either explicit or informational access was prevented by a metallic shield. Again, rats clearly differentiated between these informational contexts.

These experiments were motivated by comparing and contrasting two types of computational models: associative theories which are tied to covariation information, and causal model theory which provides a framework for translating covariation information into deeper causal representations. The evidence reviewed provides evidence that, at least in some cases, the behavior of rats is better explained by causal model theory than by associative accounts. The evidence further suggests that rats, at least to some extent, go beyond the information given to form causal model representations.

Of course, we cannot yet claim based on our current data that rats' causal knowledge has the same sophistication as the causal knowledge of humans (Penn et al., 2008). They surely lack knowledge about mechanisms and lack an understanding of the abstract concept of causality, including notions of relations about relations and related analogical cognition. Nevertheless we have shown that rats' causal reasoning goes beyond simple associative theories, and embody many aspects of causality that are crucial components of causal representations (e.g., causal directionality; interventional inferences).

We certainly do not claim that rats have a metacognitive understanding of their causal knowledge; probably much of human behavior also occurs in the absence of causal self knowledge. Our studies test between theories on the computational level; they do not test how the computational accounts are actually represented in terms of mechanisms. Thus, when we criticize associative theories, we discuss them as computational theories which deny that organisms go beyond covariation information in causal reasoning, not as accounts of possible neural mechanisms.

It may well be that our favored account, causal model theory, will eventually be implemented or subsumed within a complex theory that uses associations as the basic building block. There are many examples of processes that operate on the associative networks of the nervous system, but that result in complex cognition. For example, vertebrates such as humans, rats, and pigeons often engage in pattern completion (such as second-order conditioning, sensory preconditioning, transitive inference, sequence learning, etc.) when some elements of a pattern are missing. Pattern completion has been found for spatial (Blaisdell & Cook, 2005; Chamizo, Roderigo, & Mackintosh, 2006; Sawa, Leising, & Blaisdell, 2005), temporal (Arcediano, Escobar, & Miller, 2003; Leising, Sawa, & Blaisdell, 2007), as well as in Pavlovian conditioning (Holland, 1990; Holland & Wheeler, 2009; Rudy & O'Reilly, 1999). Penn, Holyoak, and Povinelli (2008) point out that a model of relational reasoning called LISA (Learning and Inference with Schemas and Analogies, Hummel & Holyoak, 2005) “provides an existence proof that the higher-order relational capabilities of a PSS [Physical Symbol System] can, in fact, be grafted onto a neutrally plausible, distributed connectionist architecture.” (although this has been debated by others, see peer commentary on the original article.)

So far, nobody has developed a connectionist model that implements the demonstrated computational features of reasoning with causal models. There will likely be homologies, if not merely analogies, between the way the nervous system instantiates causal and other domains of knowledge, such as spatial, temporal, and equivalence relations (Blaisdell, 2009; Hawkins & Blakeslee, 2004; Urcuioli, 2008). A neurally plausible model of the mechanisms underlying causal reasoning is certainly a desideratum. We only would like to suggest that such a model needs to honor the computational constraints highlighted by causal model theory, which we have empirically validated in the research discussed in this chapter.

References

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299-352.

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin, 114*, 435-448.

Arcediano, F., Escobar, M., & Miller, R. R. (2003). Temporal integration and temporal backward associations in human and nonhuman subjects. *Learning & Behavior, 31*, 242-256.

Barnet, R. C., Cole, R. P., & Miller, R. R. (1997). Temporal integration in second-order conditioning and sensory preconditioning. *Animal Learning & Behavior, 25*, 221-233.

Barnet, R. C., Grahame, N. J., & Miller, R. R. (1991). Comparing the magnitudes of second-order conditioning and sensory preconditioning effects. *Bulletin of the Psychonomic Society, 29*, 133-135.

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. In I. H.L. Roediger (Ed.), *Cognitive Psychology of Memory. Vol. 1 of Learning and Memory: A Comprehensive Reference* (pp. 173-195). Oxford: Elsevier.

Blaisdell, A. P. (2009). The role of associative processes in spatial, temporal, and causal cognition. In S. Watanabe, A. P. Blaisdell, L. Huber & A. Young (Eds.), *Rational animals, irrational humans* (pp. 153-172). Tokyo: Keio University Press.

Blaisdell, A., & Cook, R. G. (2005). Integration of spatial maps in pigeons. *Animal Cognition, 8*, 7-16.

Blaisdell, A. P., Leising, K. J., Stahlman, W. D., & Waldmann, M. R. (2009). Rats distinguish between absence of events and lack of information in sensory preconditioning. *International Journal of Comparative Psychology, 22*, 1-18.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*(5763), 1020-1022.

Bonawitz, E. B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A. N., Woodward, J., & Schulz, L. E. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition, 115*, 104-117.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114*, 80-99.

Bouton, M. E., (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science, 3*, 49-53.

Brogden, W. J., (1939). Sensory pre-conditioning. *Journal of Experimental Psychology, 25*, 323-332.

Buehner, M. J., & Cheng, P. W. (2005). Causal learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 143-168). Cambridge: Cambridge University Press.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: a test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 1119-1140.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Call, J. (2006). Descartes' two errors: Reason and reflection in the great apes. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 219-234). Oxford: Oxford University Press.

Castro, L., & Wasserman, E. A. (2009). Rats and infants as propositional reasoners: A plausible possibility? *Behavioral and Brain Sciences, 32*, 203-204.

Chamizo, V. D., Roderigo, T., & Mackintosh, N. J. (2006). Spatial integration with rats. *Learning & Behavior, 34*, 348-354.

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

Clayton, N., & Dickinson, A. (2006). Rational rats. *Nature Neuroscience, 9*, 472-474.

Clayton, N., Emery, N., & Dickinson, A. (2006). The rationality of animal memory: Complex caching strategies of western scrub jays. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* Oxford: Oxford University Press.

Cole, R. P., Barnet, R. C., & Miller, R. R. (1995). Temporal encoding in trace conditioning. *Animal Learning & Behavior, 23*, 144-153.

Danks, D. (2008). Rational analyses, instrumentalism, and implementations. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 59-75). Oxford: Oxford University Press.

Dickinson, A., & Balleine, B. (1993). Actions and responses: The dual psychology of behavior. In N. Eilan & R. A. McCarthy (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 277-293). Malden, MA: Blackwell Publishers Inc.

Dowe, P. (2000). *Physical causation*. Cambridge, UK: Cambridge University Press.

Dwyer, D. M., Starns, J., & Honey, R. C. (2009). "Causal reasoning" in rats: A reappraisal. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 578-586.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107*, 289-344.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: The MIT Press.

Glymour, C., & Stalker, D. (1980). *Theory and Evidence*. Princton University Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*(1), 3-32.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Science, 8*, 371-377,

Gopnik, A., & Schulz, L. (2007). *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354-384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116,* 661-716.

Hagmayer, Y., & Sloman, S. A. (2009). Decision makers conceive of their choices as interventions. *Journal of Experimental Psychology: General, 138*, 22-38.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition, 30*, 1128-1137.

Hagmayer, Y., & Waldmann, M. R. (2007). Inferences about unobserved causes in human contingency learning. *Quarterly Journal of Experimental Psychology, 60*, 330-355.

Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Ecco (HarperCollins).

Hawkins, J. & Blakeslee, S. (2004). *On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines*. New York: Henry Holt and Company, LLC.

Holland, P. C. (1990). Event representation in Pavlovian conditioning: image and action. *Cognition, 37*, 105-131.

Holland, P. C., & Wheeler, D. S. (2009). Representation-mediated food aversions. In S. Reilly & T. R. Schachtman (Eds.), *Conditioned taste aversion: Behavioral and neural processes* (pp. 196-225). New York: Oxford University Press.

Hume, D. (1748/1977). *An enquiry concerning human understanding*. Indianapolis: Hackett Publishing Company.

Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally plausible cognitive architecture: An overview of the LISA project. *Current Directions in Psychological Science, 14*, 153-157.

Kacelnik, A. (2006). Meanings of rationality. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 87-106). Oxford: Oxford University Press.

Killeen, P. R. (1978). Superstition: A matter of bias, not detectability. *Science, 199*, 88-90.

Killeen, P. R. (1981). Learning as causal inference. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 89-112). Cambridge, MA: Ballinger.

Krechevsky, I. (1932). 'Hypotheses' in rats. *Psychological Review, 39*, 516-532.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science, 16*, 678-683.

Kushnir, T., Gopnik, A., Lucas, C., & Schulz, L. E. (2010). Inferring hidden causal structure. *Cognitive Science, 34*, 148-160.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 856-876.

Lagnado, D. A., Waldmann, M. A., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation. Cues to causal structure. In A. Gopnik, A., & L. E. Schulz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154-172). Oxford University Press.

Leising, K. J., Sawa, K., & Blaisdell, A. P. (2007). Temporal integration in Pavlovian appetitive conditioning in rats. *Learning & Behavior, 35*, 11-18.

Leising, K. J., Wong, J., Waldmann, M. R., & Blaisdell, A. P. (2008). The special status of actions in causal reasoning in rats. *Journal of Experimental Psychology: General, 137*, 514-527.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*, 265-288.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology, 40*, 87-137.

Liljeholm, M., & Cheng, P. W. (2007). When is a cause the "same"? Coherent generalization across contexts. *Psychological Science, 18*, 1014-1021.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review, 15*, 75-80.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & Cognition, 37*, 249-264.

Michotte, A. (1946/1963). *The perception of causality*. London: Methuen.

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

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Science, 5*, 309-315.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686-713.

Munakata, Y., & Yerys, B. E. (2001). All together now: when dissociations between knowledge and action disappear. *Psychological Science, 12*, 335-337.

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford Univ. Press.

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

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, U.K.: Cambridge University Press.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behavioural Brain Science, 31*, 109-130; discussion 130-178.

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

Povinelli, D. J. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works*. Oxford/New York: Oxford University Press.

Quine, W. V. (1960). *Word and object*. The MIT Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rudy, J. W., & O'Reilly, R. C. (1999). Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. *Behavioral Neuroscience, 113*, 867-880.

Sawa, K. (2009). Predictive behavior and causal learning in animals and humans. *Japanese Psychological Research, 51*, 222-233.

Sawa, K., Leising, K. J., & Blaisdell, A. P. (2005). Sensory preconditioning in spatial learning using a touch screen task in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 368-375.

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43*, 149-158.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*, 322-332.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 229-261). New York: Academic Press.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. Oxford: Oxford University Press.

Sloman, S., & Fernbach, P. M. (2008). The value of rational analysis: An assessment of causal reasoning and learning. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Perspectives for Bayesian cognitive science* (pp. 485-500). Oxford: Oxford University Press.

Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences, 10*, 407-412.

Sloman, S. A., & Lagnado, D. A. (2004). Causal invariance in reasoning and learning. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 287–325). San Diego: Elsevier Science.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science, 29*, 5-39.

Sneed, J. D. (1971). *The logical structure of mathematical physics*. Dordrecht, D. Reidel Publishing Company.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science.Special Issue: 2002 Rumelhart Prize Special Issue Honoring Richard Shiffrin, 27*, 453-489.

Taubes, G. (2007). *Good calories, bad calories: Challenging the conventional wisdom on diet, weight control, and disease*. New York: Alfred & Knopf.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*, 309-318.

Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review, 42*, 43-77.

Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University Press.

Urcuioli, P. J. (2008). Associative symmetry, antisymmetry, and a theory of pigeons' equivalence-class formation. *Journal of the Experimental Analysis of Behavior, 90*, 257-282.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 34: Causal learning* (pp. 47-88). San Diego: Academic Press.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 53-76.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8*, 600-608.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for rational models of cognition* (pp. 453-484). Oxford: Oxford University Press.

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science, 18*, 247-253.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*, 27-58.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 216-227.

Waldmann, M. R., & Hagmayer, Y. (2006). Categories and causality: The neglected direction. *Cognitive Psychology, 53,* 27-58.

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models of learning and reasoning. *Current Directions in Psychological Science, 15*, 307-311.

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.

Waldmann, M. R., Schmid, M., Wong, J., & Blaisdell, A. P. (2010). Rats distinguish between absence of events and lack of evidence in contingency learning. Unpublished Manuscript.

Wasserman, E. A., (1990). Detecting response-outcome relations: Toward an understanding of the causal texture of the environment. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation.* San Diego: Academic Press. Pp. 27-82.

White, P. A. (2006). The role of activity in visual impressions of causality. *Acta Psychologica, 123*, 166-185.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science, 10*, 92-97.

Yin, H., Barnet, R. C., & Miller, R. R. (1994). Second-order conditioning and Pavlovian conditioned inhibition: Operational similarities and differences. *Journal of Experimental Psychology: Animal Behavior Processes, 20*, 419-428.

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

Young, M. E., Beckmann, J. S., & Wasserman, E. A. (2006). Pigeons’ discrimination of Michotte’s launching effect. *Journal of the Experimental Analysis of Behavior, 86*, 223-237.

Figure Captions

Figure 1. Observing an effect (left) versus intervening in an effect (right) of a common cause: While an observation of an effect allows inferring the presence of its cause, an intervention in the same variable renders this variable independent of its cause. See text for details.

Figure 2. Photograph of a rodent conditioning chamber used in the research from the Blaisdell lab discussed in this chapter.

Figure 3. Left panel: Causal model presented to rats in Blaisdell et al. (2006, Experiment 1). Center panel: Each causal link was presented separately ('→' signifies temporal order, ':' signifies simultaneous presentation). Right panel: Test trials presented either the alternative effect of the cause of Food (Tone), the second cause of Food (Click), or these two events as a causal outcome of lever presses (Click and Tone were counterbalanced). Rats' expectations of the presence of Food were assessed by measuring their search behavior (nose poking). See text for further details.

Figure 4. Results of Blaisdell et al. (2006, Experiment 1). Rats either observed Tone and intervened in Click, or observed Click and intervened in Tone (Tone and Click were counterbalanced). Error bars represent standard errors of the mean.

Figure 5. Predictions of causal model theory for each test condition of Blaisdell et al. (2006), Experiment 1: Common Cause Intervene (top-left panel), Common Cause Observe (bottom-left panel), Direct Cause Intervene (top-right panel), and Direct Cause Observe (bottom-right panel). Graph surgery is predicted only in condition Common Cause Intervene depicted as the deletion of the arrow from the Light to the Tone resulting from the lever press→Tone contingency at test. Acknowledgement to Bernard Balleine for permission to use the cartoon rat.

Figure 6. Correlations between Mean trial Nose Pokes and Mean trial Lever Presses from Intervene test for Groups Common Cause (diamond symbols) and Direct Cause (square symbols) from Blaisdell et al. (2006), Experiment 1 (top panel) and Dwyer et al. (2009), Experiment 1 (bottom panel).

Figure 7. Results of Test Days 1 and 2 from Experiment 1 of Dwyer et al. (2009) that serve as a direct replication of the design used by Blaisdell et al. (2006).

Figure 8. Predictions of causal model theory for each test condition of Blaisdell et al. (2006), Experiment 2a: Common Cause Intervene (left panel) and Causal Chain Intervene (right panel). Graph surgery is predicted only in condition Common Cause Intervene depicted as the deletion of the arrow from the Light to the Tone

resulting from the lever press→Tone contingency at test. Acknowledgement to Bernard Balleine for permission to use the cartoon rat.

Figure 9. Left panel: Mean nose pokes during the Light during Phase 2 Light→Sucrose pairings. Right panel: Mean nose pokes during the Tone at test. Error bars denote standard errors of the means. From Experiment 2a of Blaisdell et al., 2006. (Adapted with permission of *Science*).

Figure 10. Predictions of causal model theory for each test condition of Leising et al. (2008), Experiments 1 and 2: Common Cause Intervene (top-left panel) and Exogenous-Cue Intervene (top-right panel). Graph surgery is predicted only in condition Common Cause Intervene depicted as the deletion of the arrow from the Light to the Tone resulting from the lever press→Tone contingency at test. It is questionable whether a Click will produce graph surgery (depicted by the '?' in place of the arrow between the Light and the Tone). Bottom panel: Experimental design of Leising et al. (2008), Experiments 1 and 2. Acknowledgement to Bernard Balleine for permission to use the cartoon rat.

Figure 11. Results of test trials from Leising et al. (2009), Experiment 1 (top panel) and Experiment 2 (bottom panel). Error bars show standard errors of the mean.

Figure 12. Left panel: Causal model presented to rats in Leising et al. (2008, Experiment 3). Center panel: Training procedure. Each causal link was presented separately ('→' signifies temporal order, ':' signifies simultaneous presentation). Right panel: Test trials presented either the alternative effect of the cause of Food (Tone), the second cause of Food (Click), or these two events as a causal outcome of lever presses (Click and Tone were counterbalanced). Rats received one of four test conditions: Intervene on Tone on Test Day 1 followed by Observe Tone on Test Day 2, Observe Tone on Test Day 1 followed by Intervene on Tone on Test Day 2, Intervene on Click on Test Day 1 followed by Observe Click on Test Day 2, Observe Click on Test Day 1 followed by Intervene on Click on Test Day 2. Rats' expectations of the presence of Food were assessed by measuring their search behavior (nose poking). See text for further details.

Figure 13. Mean Nose Pokes during test trials for each test session of Leising et al. (2008) Experiment 3. Rats that intervened on the Tone on Test Day 1 observed the Tone on Test Day 2. Rats that observed the Tone on Test Day 1 intervened on the Tone on Test Day 2. Rats that Intervened on the Click on Test Day 1 observed the Click on Test Day 2. Rats that observed the Click on Test Day 1 intervened on the Click on Test Day 2. (Tone and Click were counterbalanced across conditions). Error bars represent standard errors of the mean.

Figure 14. Mean Nose Pokes during the first test trial for subjects in meta-analysis reported by Leising et al. (2008). Intervene condition includes subjects that received Common Cause training and the Intervene test. Observe condition includes subjects that received Common Cause training and the Observe test. Error bars represent standard errors of the mean.

Figure 15. Three hypothetical causal models describing the causal relationships among there variables: Smoking, Yellow Teeth, and Lung Cancer. See text for details.

Figure 16. Mean lever presses during trials with the Tone (Common Cause) and Noise (Causal Chain) in the Intervene and Observe test conditions, from Blaisdell et al. (2006) Experiment 1. Error bars represent standard errors of the mean.

Figure 17. Top panel: Mean discrimination ratios for nose poke responses during test trials with the second-order (Paired) CS and the Unpaired CS from Blaisdell et al. (2009) Experiment 2. Testing was conducted either with the Light Present or Absent. Bottom panel: Mean discrimination ratios for nose poke responses during test trials with the second-order CS with the Light Present or Absent during testing. Testing occurred either in the Same or Different context from where training took place. Error bars represent standard errors of the mean.

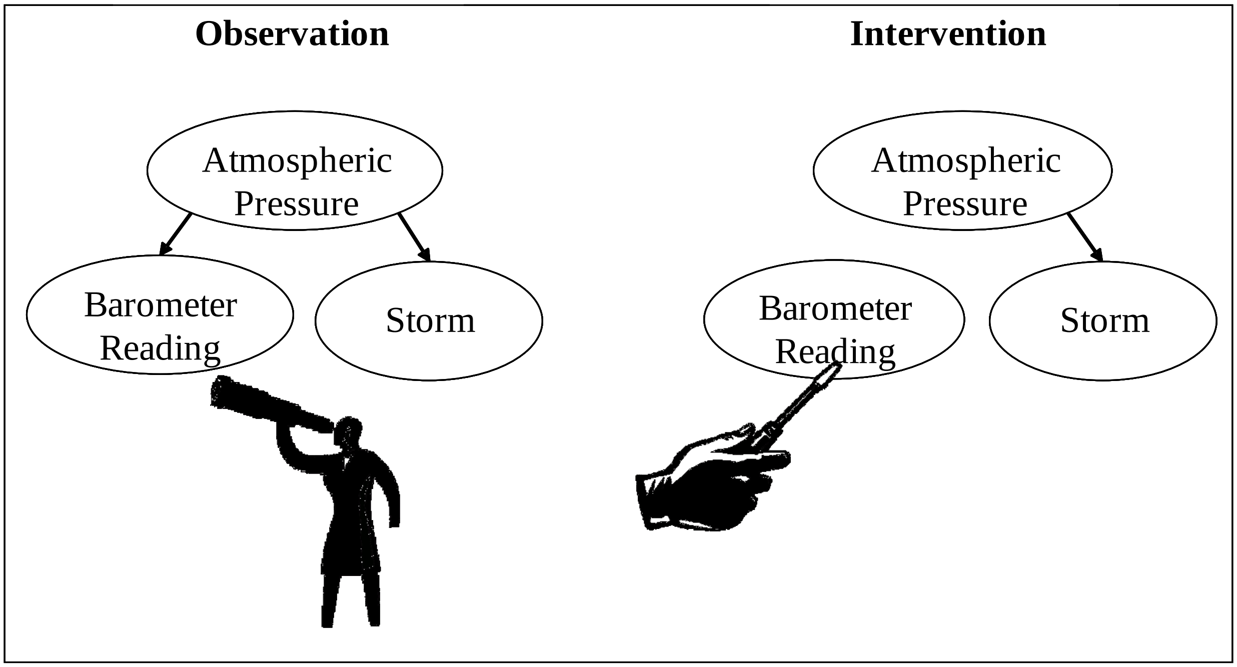

Observation
Intervention
Atmospheric Pressure
Barometer Reading
Storm
Atmospheric Pressure
Barometer Reading
Storm

Figure 1

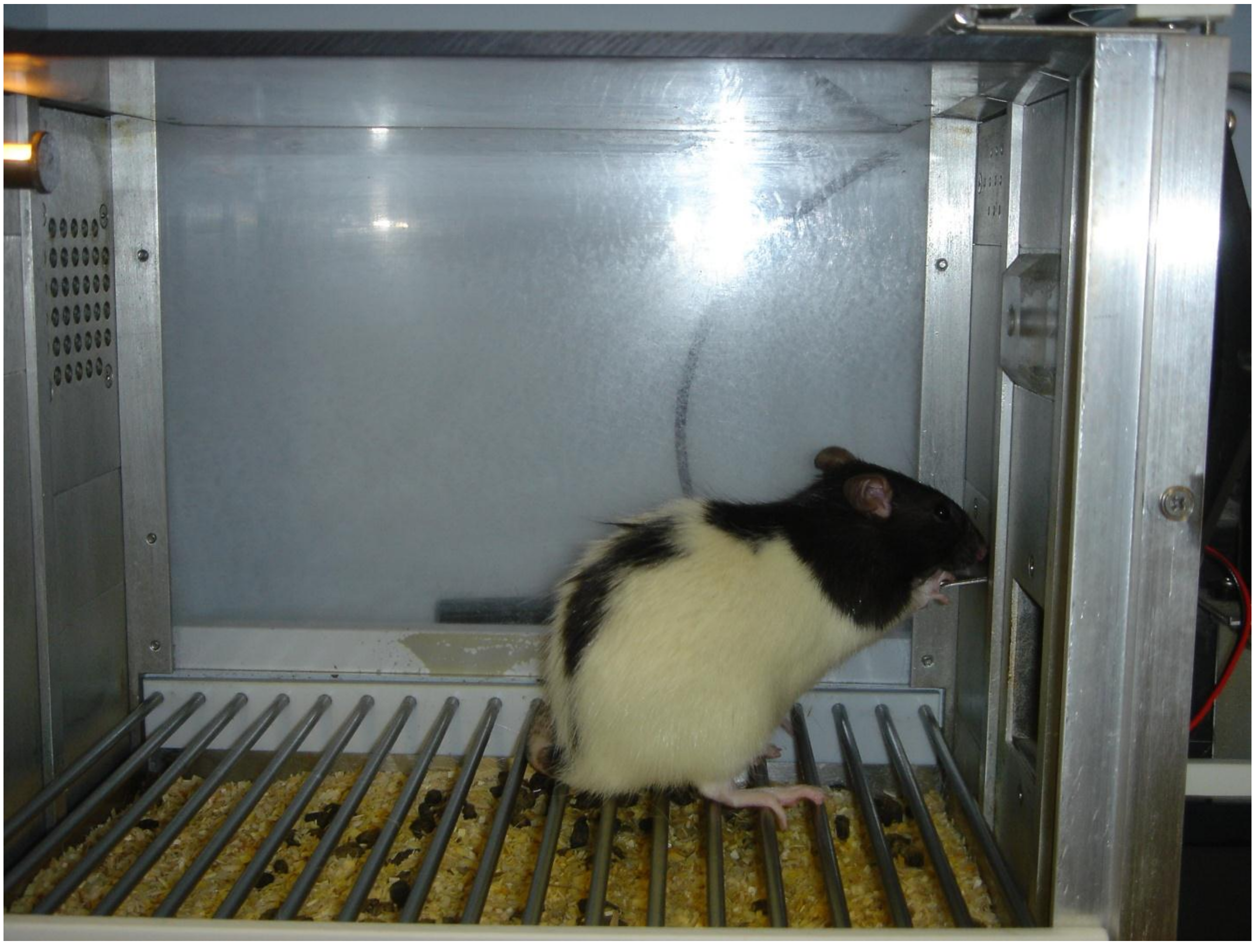


Figure 2

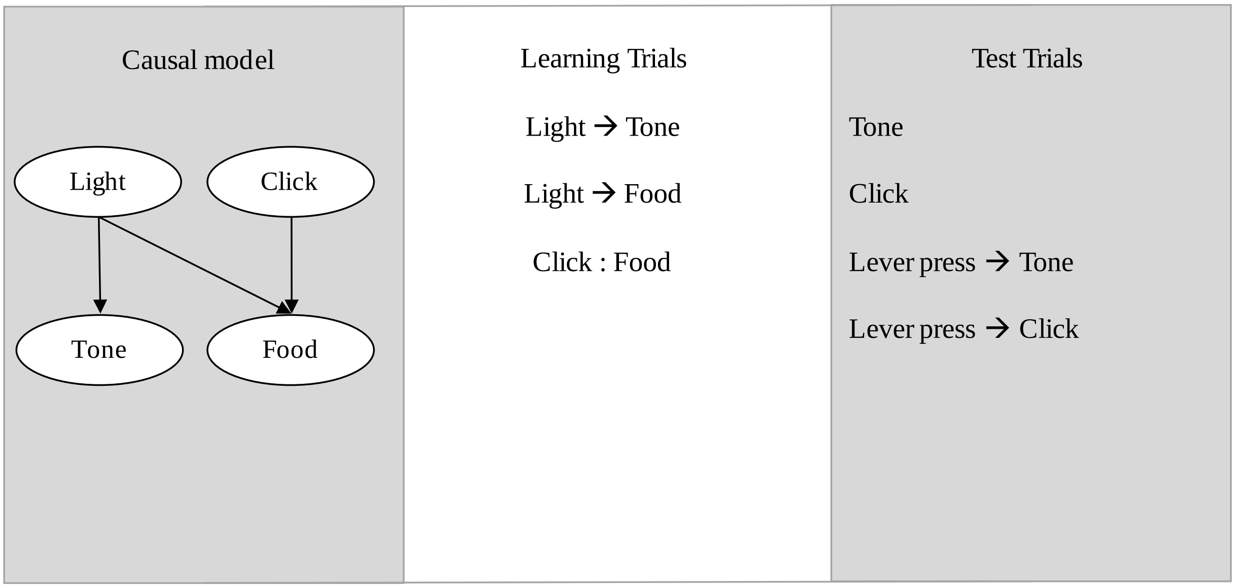

Causal model
Light
Click
Tone
Food
Learning Trials
Light → Tone
Light → Food
Click : Food
Test Trials
Tone
Click
Lever press → Tone
Lever press → Click

Figure 3

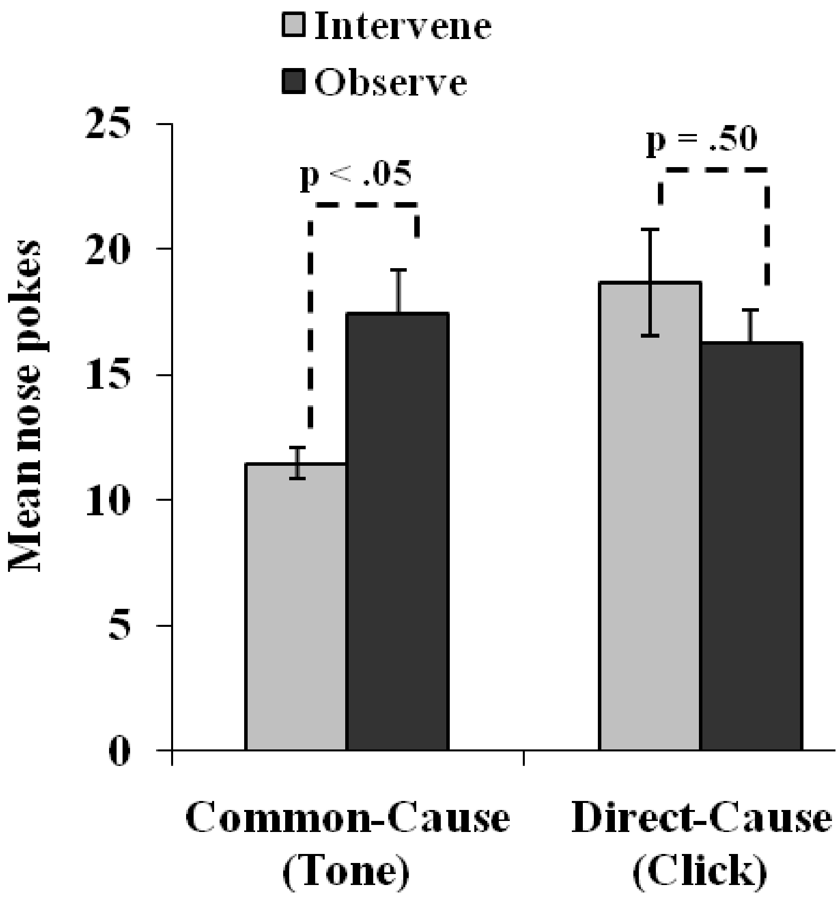

Intervene
Observe
25
20
15
10
5
0
Mean nose pokes
p < .05
p = .50
Common-Cause
(Tone)
Direct-Cause
(Click)

Figure 4

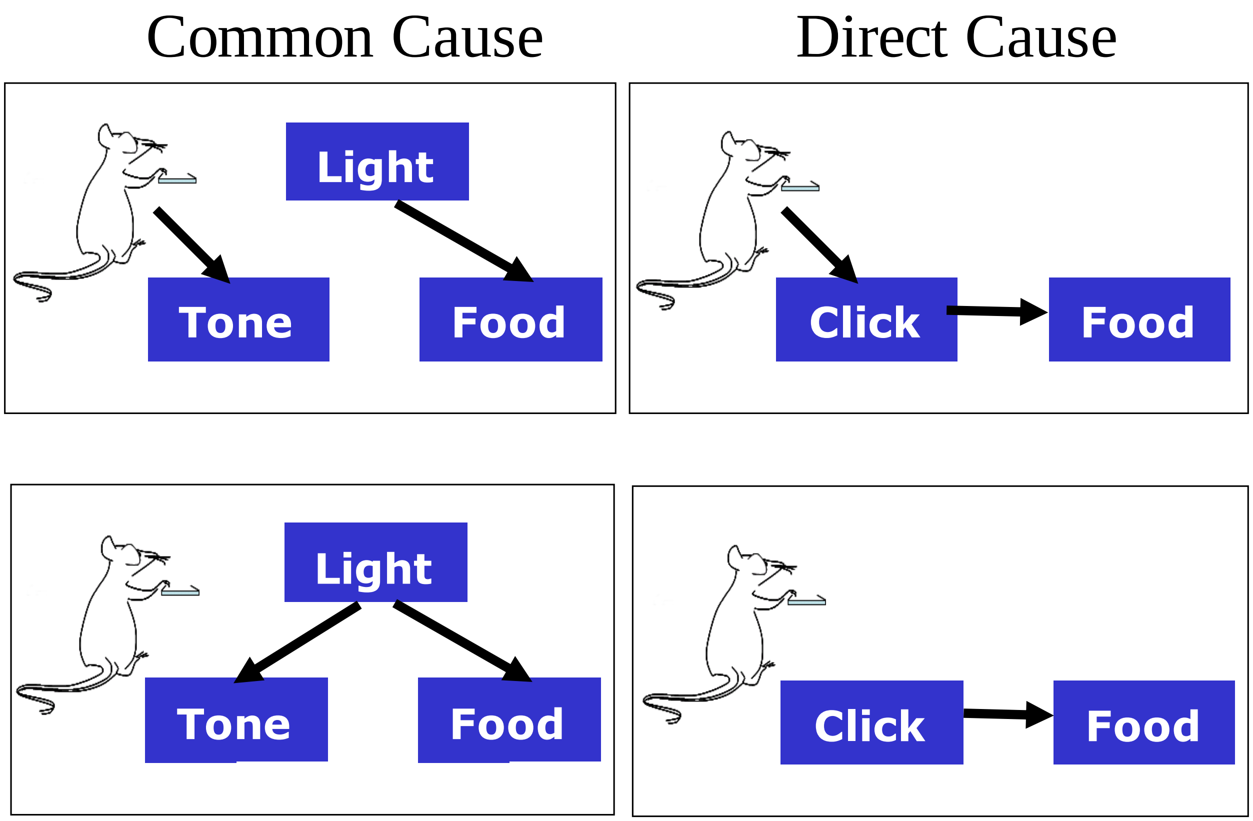
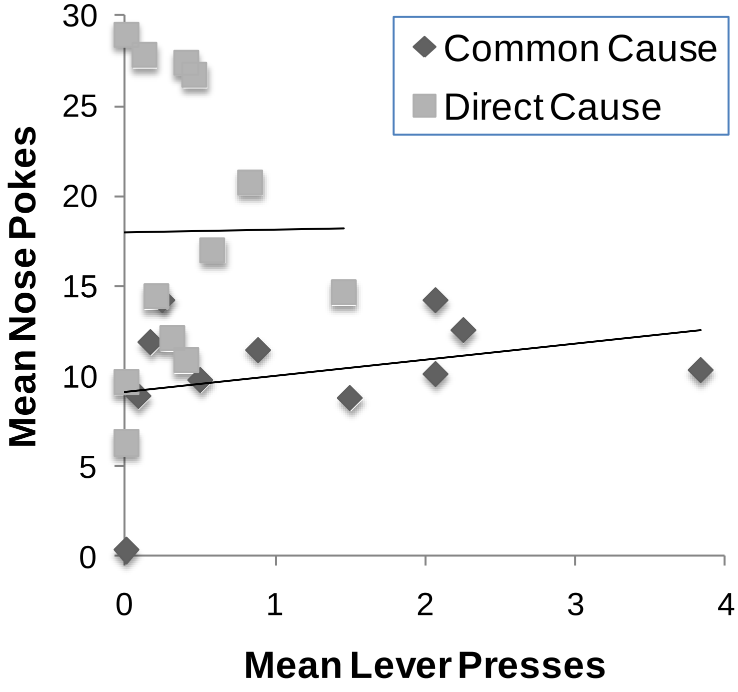

Common Cause
Direct Cause
Light
Tone
Food
Click
Food
Light
Tone
Food
Click
Food

Figure 5


30
25
20
15
10
5
0
0
1
2
3
4
Mean Nose Pokes
Mean Lever Presses
◆ Common Cause
■ Direct Cause

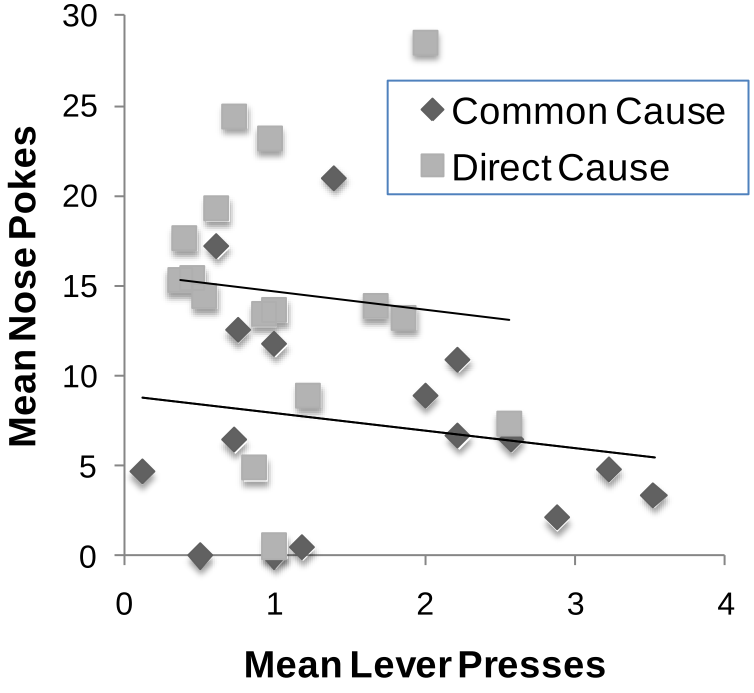

30
25
20
15
10
5
0
0
1
2
3
4
Mean Nose Pokes
Mean Lever Presses
◆ Common Cause
■ Direct Cause

Figure 6

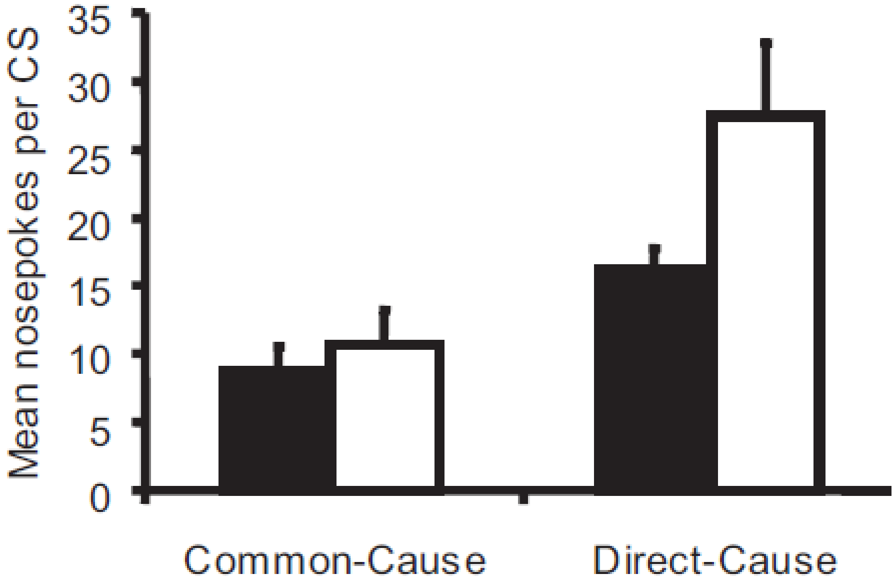

Mean nosepokes per CS
35
30
25
20
15
10
5
0
Common-Cause
Direct-Cause

Figure 7

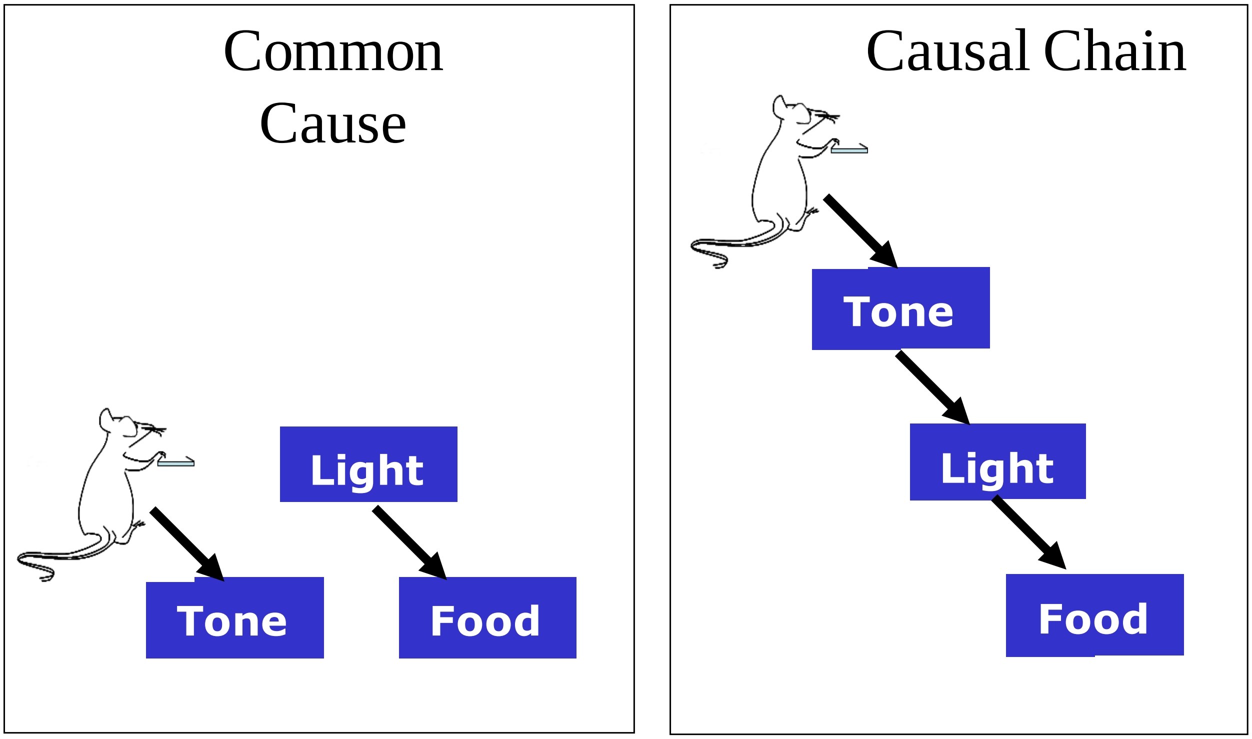


Figure 8

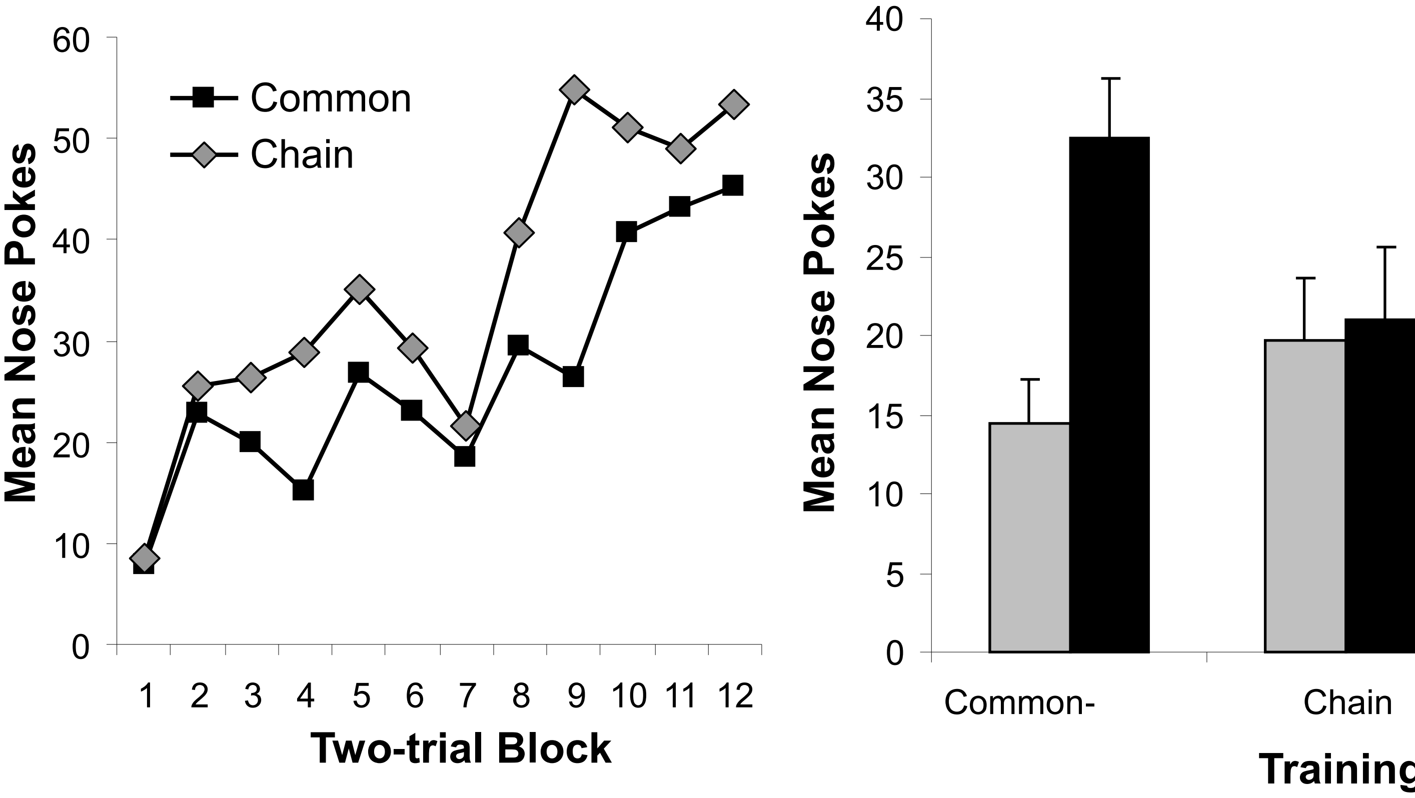
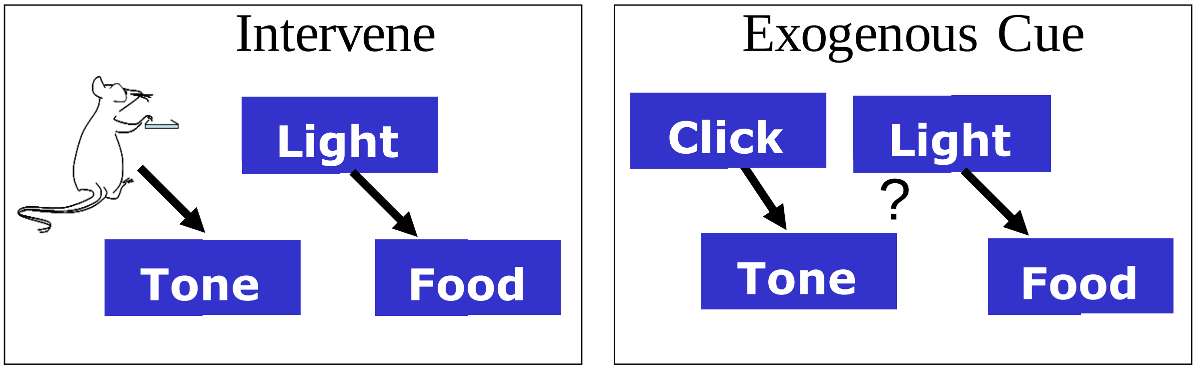

Mean Nose Pokes
60
50
40
30
20
10
0
Common
Chain
1
2
3
4
5
6
7
8
9
10
11
12
Two-trial Block
Mean Nose Pokes
40
35
30
25
20
15
10
5
0
Common-
Chain
Training

Figure 9



Group	Phase 1	Phase 2	Test
Intervene	Light→Tone	Light→Food	LP→Tone
Observe	Light→Tone	Light→Food	LP / Tone
Exogenous Cue	Light→Tone	Light→Food	LP / Click→Tone

Yoked

Figure 10

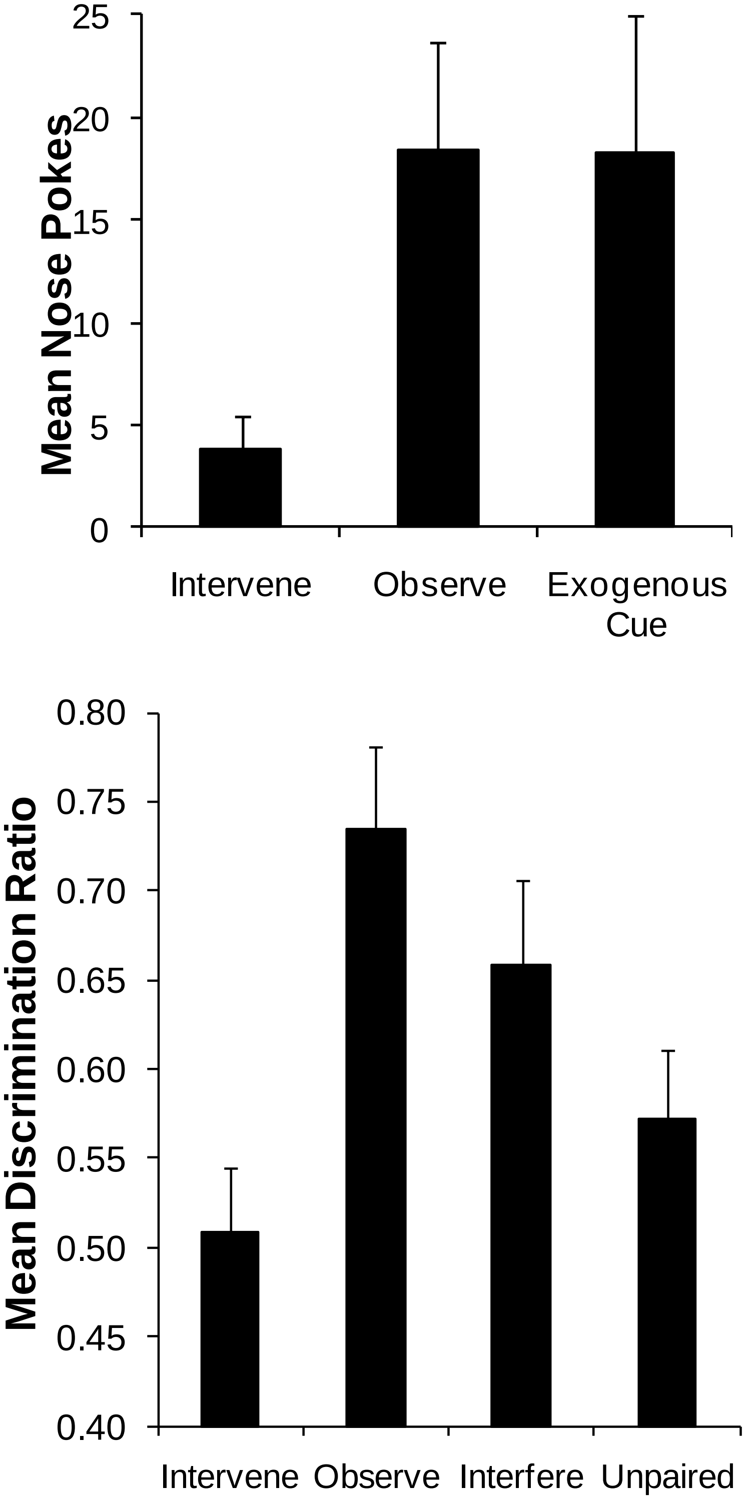
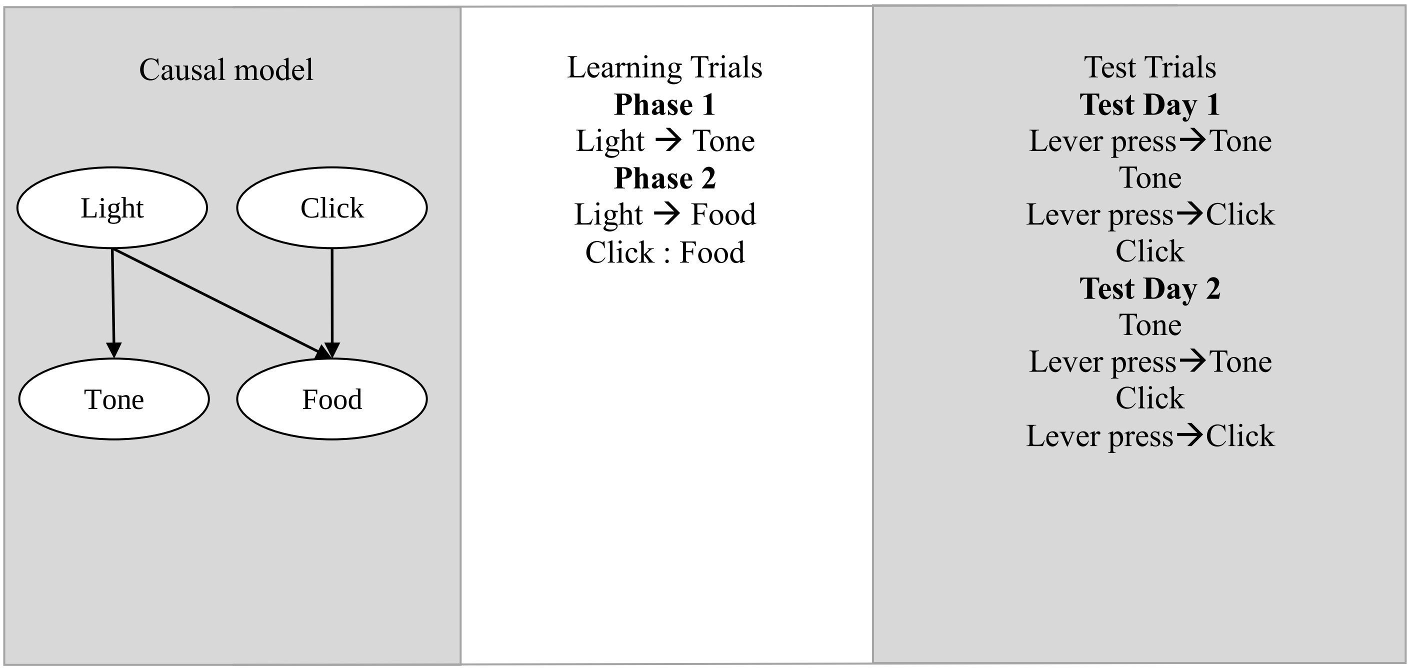

25
20
15
10
5
0
Mean Nose Pokes
Intervene
Observe
Exogenous Cue
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
Mean Discrimination Ratio
Intervene
Observe
Interfere
Unpaired

Figure 11



Learning Trials

Phase 1

Light → Tone

Phase 2

Light → Food

Click : Food

Test Trials

Test Day 1

Lever press→Tone

Tone

Lever press→Click

Click

Test Day 2

Tone

Lever press→Tone

Click

Lever press→Click

Figure 12

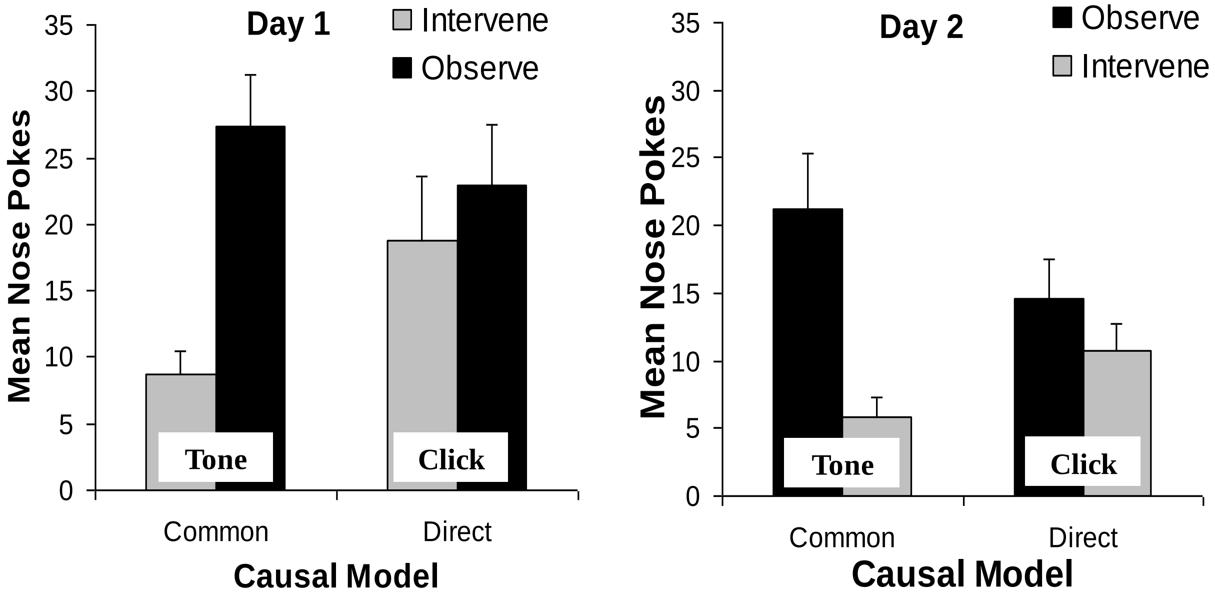

Day 1
Intervene
Observe
Mean Nose Pokes
35
30
25
20
15
10
5
0
Tone
Click
Common
Direct
Causal Model
Day 2
Observe
Intervene
Mean Nose Pokes
35
30
25
20
15
10
5
0
Tone
Click
Common
Direct
Causal Model

Figure 13

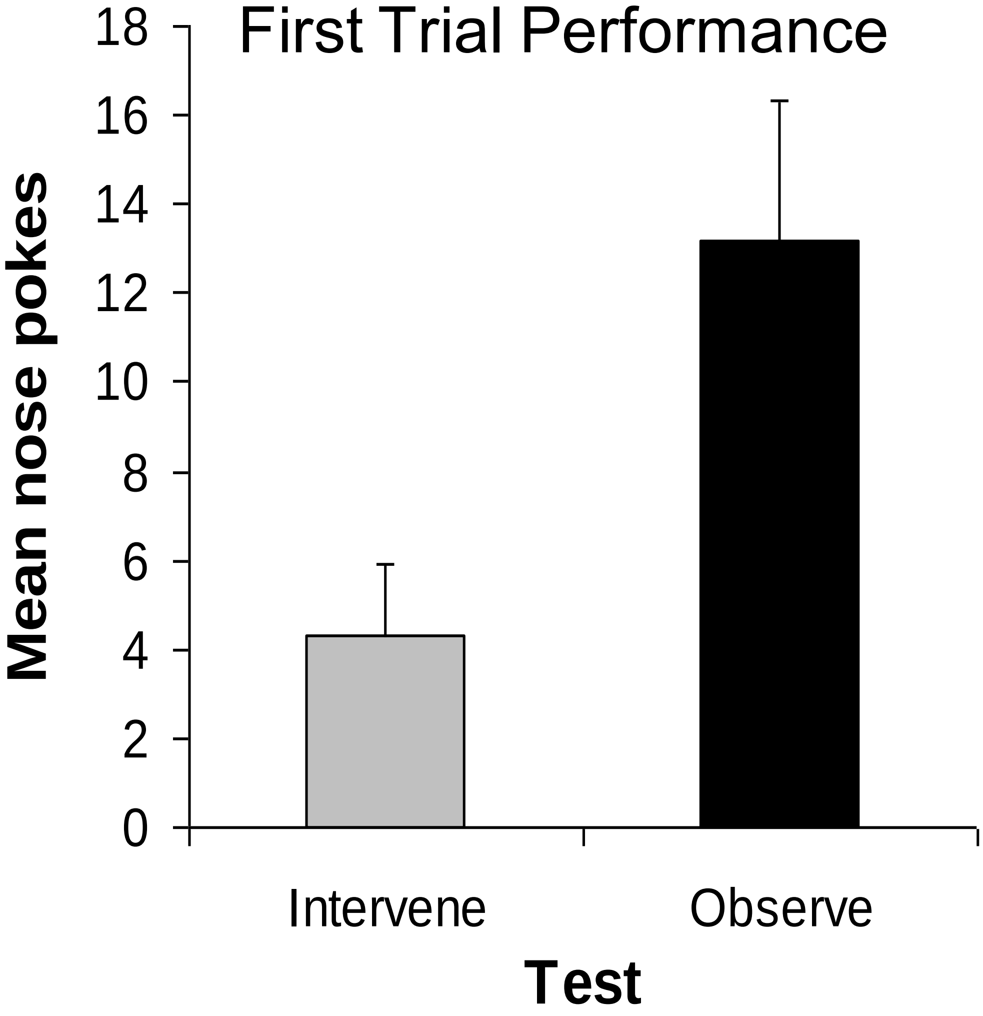

First Trial Performance
Mean nose pokes
0
2
4
6
8
10
12
14
16
18
Intervene
Observe
Test

Figure 14

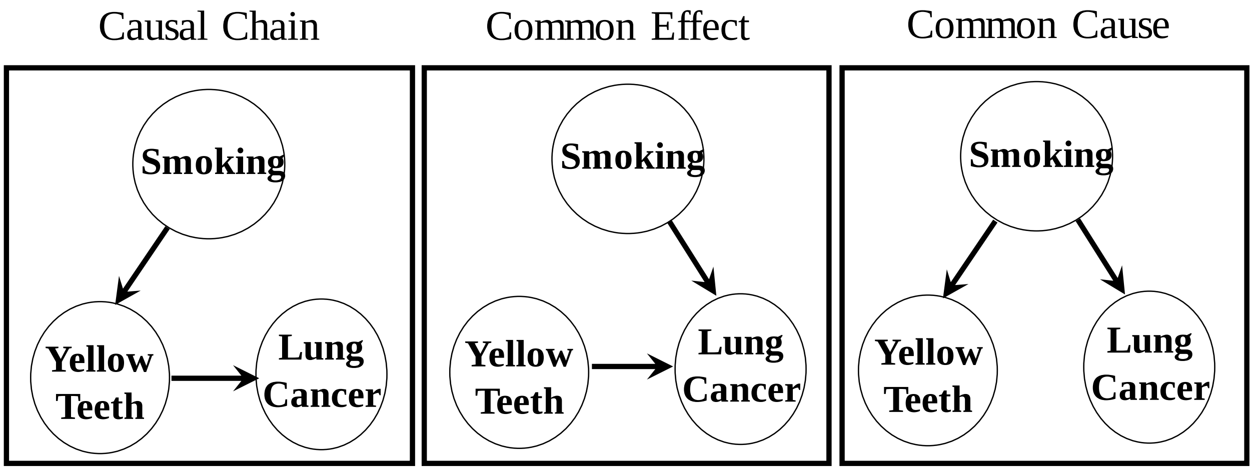

Causal Chain
Common Effect
Common Cause
Smoking
Yellow Teeth
Lung Cancer
Smoking
Yellow Teeth
Lung Cancer
Smoking
Yellow Teeth
Lung Cancer

Figure 15

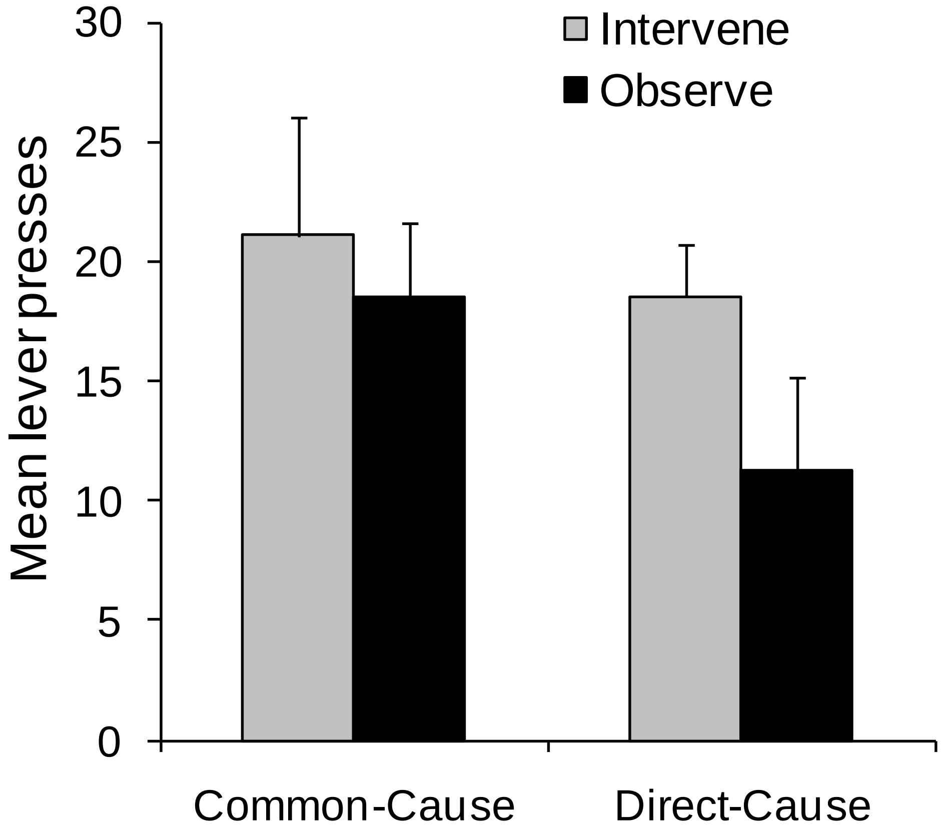

30
25
20
15
10
5
0
Mean lever presses
Intervene
Observe
Common-Cause
Direct-Cause

Figure 16

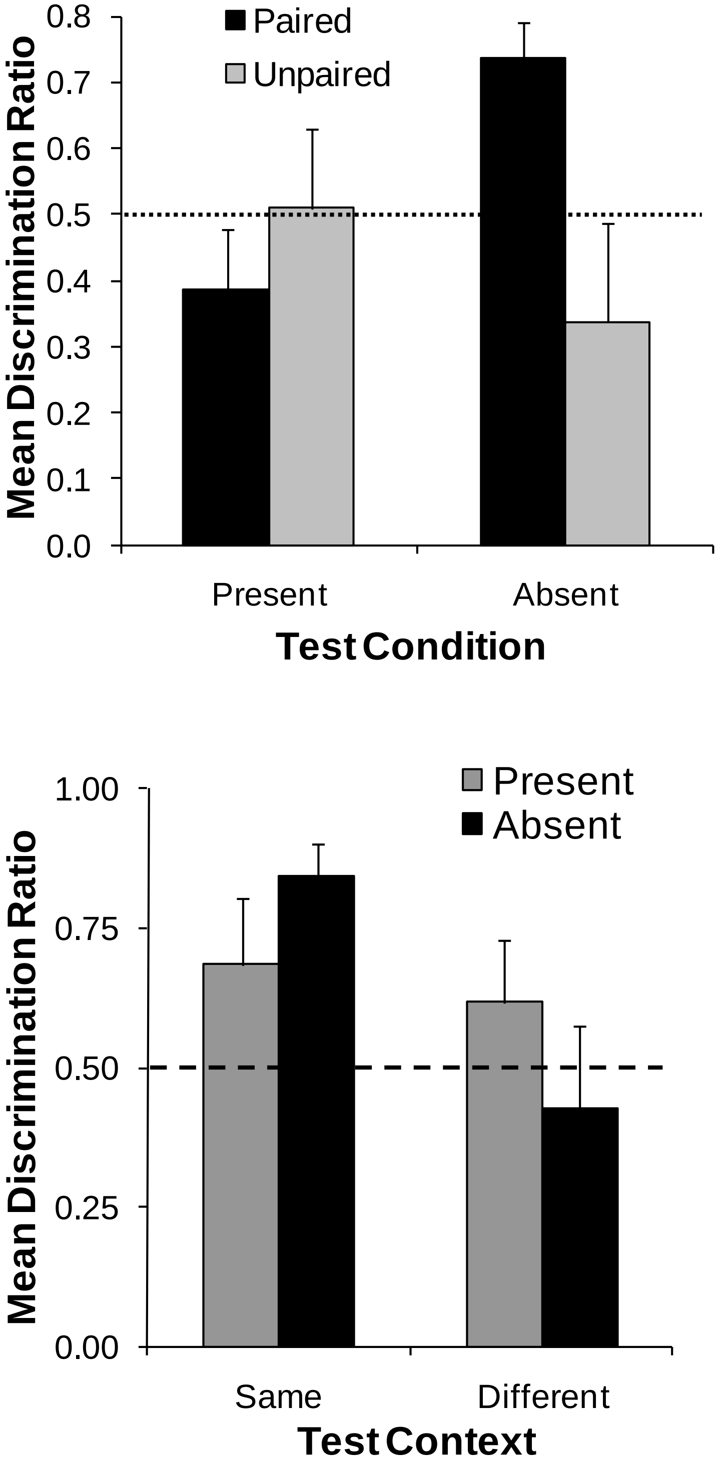

■ Paired
□ Unpaired
Mean Discrimination Ratio
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Present
Absent
Test Condition
■ Present
■ Absent
Mean Discrimination Ratio
1.00
0.75
0.50
0.25
0.00
Same
Different
Test Context

Figure 17