Journal of Experimental Psychology: General
2008, Vol. 137, No. 3, 514–527

0096-3445/08/$12.00 DOI: 10.1037/0096-3445.137.3.514

The Special Status of Actions in Causal Reasoning in Rats

Kenneth J. Leising and Jared Wong
University of California, Los Angeles

Michael R. Waldmann
University of Göttingen

Aaron P. Blaisdell
University of California, Los Angeles

A. P. Blaisdell, K. Sawa, K. J. Leising, and M. R. Waldmann (2006) reported evidence for causal reasoning in rats. After learning through Pavlovian observation that Event A (a light) was a common cause of Events X (an auditory stimulus) and F (food), rats predicted F in the test phase when they observed Event X as a cue but not when they generated X by a lever press. Whereas associative accounts predict associations between X and F regardless of whether X is observed or generated by an action, causal-model theory predicts that the intervention at test should lead to discounting of A, the regular cause of X. The authors report further tests of causal-model theory. One key prediction is that full discounting should be observed only when the alternative cause is viewed as deterministic and independent of other events, 2 hallmark features of actions but not necessarily of arbitrary events. Consequently, the authors observed discounting with only interventions but not other observable events (Experiments 1 and 2). Moreover, rats were capable of flexibly switching between observational and interventional predictions (Experiment 3). Finally, discounting occurred on the very first test trial (Meta-Analysis). These results confirm causal-model theory but refute associative accounts.

Keywords: causal model, intervention, rats, discounting, causal reasoning

Knowledge of the causal structure of the world has important functional value. Causal knowledge may be used to predict future events and to manipulate the world to achieve goals. Theories of causal knowledge fall into two categories of explanation. According to associative theories of causal cognition, representations of cause–effect relations reflect information about the contiguity and contingency between events (e.g., Allan, 1993; Shanks & Dickinson, 1987; Young, 1995). Causal-model theory, by contrast, suggests that causal learning involves going beyond covariations. According to this view, causal power and causal structures are theoretical entities that are estimated, combining cues to infer causal structure (e.g., temporal order) and statistical learning input to assess causal strength (e.g., Cheng, 1997; Waldmann, Cheng, Hagmayer, & Blaisdell, 2008; Waldmann, Hagmayer, & Blaisdell, 2006; Waldmann & Holyoak, 1992). Thus, analogous to visual perceptual processes in which retinal input is further processed to recover a representation of a three-dimensional world, causal learning recovers causal structures in the world from cues and statistical learning input.

In many instances, causal models provide more accurate representations of the world than do mere cause–effect associations derived from covariations (Waldmann et al., 2006). In particular, causal models distinguish between causal and noncausal relations between correlated events. For example, atmospheric pressure, barometer readings, and the weather are three covarying events. Though we may observe the weather to change reliably following a change in the barometer, if we understand the true underlying causal structure then we realize that both changes are due to a change in a common cause—atmospheric pressure. Whereas the covariations between the common cause, atmospheric pressure, and its two effects are causal, the relation between barometer readings and weather is noncausal. Causal-model theory allows for representing and learning such structures (see Lagnado, Waldmann, Hagmayer, & Sloman, 2007).

Another important feature of causal relations is their directionality. For example, atmospheric pressure is associated with barometer readings, as are barometer readings with atmospheric pressure; nevertheless, atmospheric pressure is the cause of barometer readings and not vice versa. This distinction is crucial for planning actions (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; Woodward, 2003).

One novel technique to test whether human or nonhuman animals have acquired representations of causal models, rather than merely associative knowledge, is to compare causal inferences when an effect of a cause is observed merely with situations in which the effect is produced by an intervention (i.e., is manipulated to occur). Sloman and Lagnado (2005) and Waldman and Hagmayer (2005) have used this technique in humans to distin-

Kenneth J. Leising, Jared Wong, and Aaron P. Blaisdell, Department of Psychology, University of California, Los Angeles; Michael R. Waldmann, Department of Psychology, University of Göttingen, Göttingen, Germany.

Support for this research was provided by National Institutes of Health Grant MH12531 (to Aaron P. Blaisdell) and Deutsche Forschungsgemeinschaft Grant WA 621/20-1 (to Michael R. Waldmann and Y. Hagmayer). This research was conducted following the relevant ethics guidelines for research with animals and was approved by the University of California, Los Angeles' Institutional Animal Care and Use Committee.

Correspondence concerning this article should be addressed to Aaron P. Blaisdell, UCLA Department of Psychology, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563. E-mail: blaisdell@psych.ucla.edu

guish between causal and associative theories (see also Hagmayer, Sloman, Lagnado, & Waldmann, 2007; Meder, Hagmayer, & Waldmann, 2008). A fictitious example of this technique would be to have participants who lack prior experience with barometers passively observe the relations between atmospheric pressure, barometers, and weather. Importantly, no instrumental learning opportunity, such as the opportunity to manipulate the mercury level of the barometer and observe its lack of effects on the weather, would be available. In the test phase, we could ask participants to make predictions about the weather based either on the observation of a change in the barometer's mercury level, or after allowing the subject to intervene on the barometer by artificially raising or lowering the mercury level. Note that in the learning phase these two events, barometer reading and weather, were positively correlated. If participants' predictions of the weather merely reflect associative relations (i.e., covariations), then they should make the same prediction regardless of whether the change in the barometer was observed or the product of an intervention. In contrast, representing these events as a common-cause model—with changes in air pressure as a common cause of changes in the mercury level of the barometer and in the weather—would imply different predictions. If we observe the barometer change, we can diagnostically infer that air pressure caused this change and thus expect the weather to change as well. If we intervene by actively manipulating the barometer, however, then we might recognize that we—and not a change in air pressure—are the cause of the change in the barometer and thus should not expect a corresponding change in the weather. Thus, if causal relations were reducible to the association derived from the observed correlation between the barometer and the weather, we would incorrectly infer a causal relation between the barometer and the weather and might be duped into an attempt to manipulate the barometer so as to alter the weather.

Within causal-model theory, interventions represent external causes, which typically are independent of the other events in the causal system. Whenever interventions deterministically and independently fix states of a variable in the causal system, this variable should be viewed as fully explained by the intervention rather than its usual cause. Thus, interventions should lead to full discounting of the previous cause (Waldmann et al., 2008). One way to express this graphically is to delete the causal arrow between the previous cause and the manipulated event, which means that during the intervention these two events become statistically independent (see Figure 1, right side; Pearl, 2000; Spirtes et al., 1993).

Notably, associative theory can account for the difference between observational and interventional predictions if both prior observational and instrumental learning opportunities were provided. What standard associative theories cannot capture, however, are the inferences drawn about interventions in the absence of prior knowledge of the instrumental contingency between the intervening action (e.g., tampering with a barometer) and its effect on observed correlated relations (e.g., barometer and weather).

Adult humans have been shown to naturally represent the distinction between observation and intervention without requiring prior instrumental learning (e.g., Lagnado & Sloman, 2004; Meder et al., 2008; Waldmann & Hagmayer, 2005). Even children appear capable of this distinction (Gopnik et al., 2004). While it may be no surprise that human adults and children are capable of distinguishing between observation and intervention, it has been contentious to hypothesize this capacity in nonhuman animals. Most of the available evidence indicates that both primates (Limongelli, Boysen, & Visalberghi, 1995; Povinelli, 2000; Tomasello & Call, 1997; Visalberghi & Limongelli, 1994) and corvids (e.g., crows, jays, and rooks; see Seed, Tebbich, Emery, & Clayton, 2006) lack an understanding of the physical causal mechanisms underlying tool use (see discussion by Penn & Povinelli, 2007).

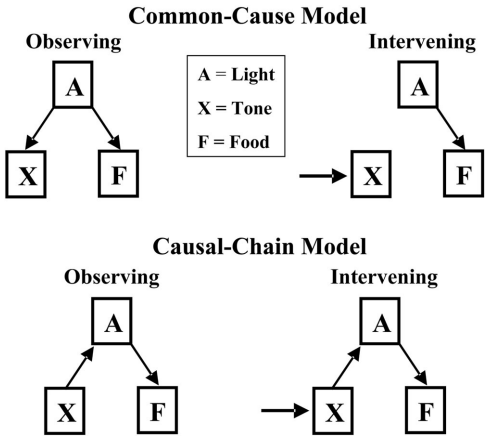

Figure 1. Common-cause (top) and causal-chain (bottom) models from Blaisdell et al. (2006; Experiment 2). Arrows indicate directionality from cause to effect. The left side diagrams the observed causal relations. The right side is modified under the assumption of an independent, deterministic intervention in X that leads to full discounting of A (i.e., statistical independence between A and X) in the common-cause model but not in the causal-chain model.

Nevertheless, Blaisdell, Sawa, Leising, and Waldmann (2006) demonstrated that rats are capable of distinguishing observational and interventional inferences and can derive correct predictions for novel interventions after purely observational learning. In their Experiments 1 and 2a, they gave rats Pavlovian pairings of Stimulus A (a light) followed by Stimulus X (a tone; A→X) and pairings of Stimulus A with food (A→F) in conventional conditioning chambers. Thus, rats observed the A–X and the A–F correlations, which may have allowed them to represent the causal structure of these correlations as Stimulus A serving as a common cause for Stimuli X and F. This causal structure is represented as X←A→F in causal graphical notation (Pearl, 2000; see Figure 1 top panel). Rats then received one of two test conditions. In the intervene test condition, a novel lever was inserted into the chamber and presses on the lever caused the onset of Stimulus X. In the observe test condition, Stimulus X was presented to the rat independently of its pressing the lever. If rats had integrated the A→X and A→F trials into a common-cause model with A as a common cause of both X and F, then based on causal-model theory, when rats observe X they should reason backward that A has occurred and thus should expect other effects of A, in particular F. In contrast, when rats intervene on X they should conclude that they, not A, are the cause of X. Thus, in this case the presence of X does not make A more likely than usual, which in turn implies that F is not more likely either. Specifically, rats that merely observed X at test (see observe test condition in Figure 1, top left section) should attribute the observation of X to A and predict that F should also be present because both events are caused by A. Rats that inter-

vened to produce X via a lever press (LP; see intervene test condition in top right section of Figure 1), however, should attribute X to their own intervention and therefore discount A and hence should not expect F to be present. Thus, causal-model theory leads to the prediction that rats in the observe test condition should exhibit a higher expectation of F during X than should rats in the intervene test condition.

Blaisdell et al. (2006) measured expectation of F by monitoring feeder activity in the hopper where F was delivered during training. Whenever the rat placed its nose into the food hopper (a nose poke), it disrupted an infrared photo beam projected across the entrance to the hopper. Thus, our specific predictions were that we would observe more nose poking (i.e., looking for food in the hopper) when rats merely observed X compared with when they intervened on X with an LP. The pattern of nose poke responding was consistent with these predictions. Associative theories, by contrast, predict equivalent rates of nose poking in the two test conditions. X is a predictor of A, which is a predictor of F, and therefore X is merely an indirect predictor of F. An associative framework posits that presentations of X should retrieve an expectation of F, which will lead the rat to engage in feeder activity anytime X is present at test (Pavlov, 1927; Yin, Barnet, & Miller, 1994). Because lever pressing had not previously been trained, lever pressing at test should not alter the Pavlovian conditioned response to X. Consistent with causal-model theory but not with associative theory, rats showed a greater expectation of F after an observation of X than after an intervention in X.

To further support the predictions of causal-model theory, Blaisdell et al. (2006) trained a second group of rats on an X→A→F causal chain by showing them correlations between X and A and between A and F. If the rats represented the correlations as a causal chain, then they should treat X as a cause of A, which in turn should be a treated as a cause of F (see lower panel of Figure 1). On a subsequent test with X alone, it should not matter whether rats merely observed X or intervened on the LP to produce X; in either case rats should expect F to occur and thus should nose poke into the food hopper. This is exactly what Blaisdell et al. observed. For rats that learned a causal chain, the LP intervention at test did not disrupt nose poking relative to the observe test. These results support the view that the rats understood the causal relationship between their action and an outcome.

Although the effect of interventions on causal expectancy reported above provides strong evidence that rats appear to reason in a sophisticated manner about causal relations, an alternative explanation for this effect invokes the influence of associative processes. There is a long history in associative learning that documents how more recently acquired associations can interfere with the retrieval and expression of previously acquired associations (Bouton, 1993; Matute & Pineño, 1998; Underwood, 1966). Matute and Pineño (1998) trained humans on two cues (A and B) paired with the same outcome (O) by delivering observational learning trials of A→O followed by trials of B→O. If subjects had learned both associations, then they should expect O during A and B. A subsequent test found a weaker expectation of O during A than B, as evidenced by impaired responding to A on a subsequent test. Responding to A was not impaired, however, when subjects received the B→O pairings before the A→O pairings. They interpreted their results in terms of associatively based retroactive interference, in which a more recently acquired association impairs retrieval of an association learned earlier. Similar results have been found in rats (Escobar, Arcediano, & Miller, 2001; Escobar & Miller, 2003).

The results of Blaisdell et al. (2006) may be amenable to an explanation in terms of associative retroactive interference. During training, rats learned an A→X association. At test, rats in the intervene test condition learned that an LP caused X to be presented. This LP→X association acquired during testing may have retroactively interfered with the A→X association, thus explaining why these rats did not expect F (another outcome of A) as much as did rats in the observe test condition.

Whereas retroactive interference theory predicts discounting at test as the result of competition between two associative relations between arbitrary events, causal-model theory makes more specific predictions (see Waldmann et al., 2008, for formal details). According to this theory, the presence of alternative causes (e.g., LP) of a common effect (e.g., X) leads to discounting of the previously established cause (e.g., A). Discounting should be maximal when the alternative event is viewed as a deterministic cause of the effect and as statistically independent of the other cause, A. Arbitrary nonaction events can be established as independent, deterministic causes if the strong contingency and the independence of the cause with the effect becomes apparent, such as through learning (Waldmann et al., 2008; Woodward, 2003). Extensive trial information is often not available, however. For example, the rats in the studies of Blaisdell et al. (2006) experienced only a small number of pairings of lever presses and tones during the test phase and none in the learning phase. Our hypothesis is that actions, such as lever presses, are special in that they immediately create the impression of causality with few trials.

Why should there be a difference between voluntary actions and arbitrary observed events co-occurring with an effect? The main difference at least for humans is that they conceive of their actions as independent manipulations of events (i.e., free). This assumed independence of actions and other events in a causal system allows us to immediately infer strong causal relations after very few trials, indicating a contingent relation between actions and outcome because there are no possible confounds creating noncausal covariations. In contrast, we do not have an independence intuition with arbitrary observed events (e.g., a tone), which may well be confounded with other events. Thus, many more trials supporting the absence of confounding are needed to establish a strong causal relation for such events. Hence, given that the establishment of a strong causal relation is a precondition for discounting, actions should after few trials produce stronger discounting than should other exogenous events that do not emanate from the actor. For example, I could walk into a room I've never been in before, press a button on the wall, and immediately hear a noise. Alternatively, I could walk into a room I've never been in before and see a novel light followed immediately by a noise. I have a more profound sense of causality in the first case than in the second. It seems unlikely that the noise just happened to come on a split second after I happened to press the button, whereas it seems much more likely that sound and light co-occur by coincidence or are both effects triggered by a third, unobserved cause, such as a motion detector that detected my entry into the room. Underlying these different conclusions is the intuition that I, and not some other event, am the initiator of the actions and that I act freely, independent of the events in the system I am going to manipulate. This

intuition, which may be implicit, allows me to be more confident in a causal relation after one or a few observations when actions and outcomes are observed as compared with when two arbitrary events co-occur.

Humans seem to conceive of their own actions as free and independent, which has (at least indirectly) been demonstrated in a number of studies (Gopnik et al., 2004; Lagnado & Sloman, 2004; Waldmann & Hagmayer, 2005; Wegner, 2002). One important goal of our three experiments was to investigate whether for rats actions also are special in that they tend to treat their actions as more likely to be causal than an arbitrary external event. Of course, such a demonstration does not necessarily imply that rats are in any way aware of why actions have a special status, but it may demonstrate that rats make similar distinctions as humans (see also General Discussion).

We test whether actions are treated by rats as special in Experiments 1 and 2 by training them on the common-cause model as in Blaisdell et al. (2006) and then, in the test phase, compare the effectiveness of an external auditory stimulus presented prior to X (a tone) at test to produce causal discounting relative to the effectiveness of an LP intervention that generates X. Experiment 3 assesses the flexibility of causal inferences regarding observations and interventions.

Experiment 1

In this experiment, we trained rats on a common-cause model like that described in Blaisdell et al. (2006). Rats first observed the correlation between two audiovisual cues, A and X, followed by a phase in which they observed the correlation between A and F (food). Subjects then received one of three test conditions. The first two test conditions replicated the observe and intervene test conditions of Blaisdell et al. (2006). In the third, exogenous cue, test condition, rats observed an auditory Stimulus B (a clicker) followed immediately by X. The design of Experiment 1 is presented in Table 1. We measured the rats' expectation of F by using the nose poke measure described in the introduction. If a rat's actions impart a more profound sense of causality than does B, then we should see more discounting in the intervention condition than in the exogenous cue condition. This prediction is based on the assumption that actions, due to their characteristic of independence, should create an immediate impression of causality with few trials, whereas the number of trials in the test phase does not suffice to establish causality for an arbitrary observed event.

Table 1
Design of Experiment 1

Group	Phase 1	Phase 2	Test
Intervene	A→X	A→F	LP→X
Observe-1	A→X	A→F	X
Exogenous cue	A→X	A→F	B→X
Observe-2	A→X	A→F	→X

Note. An arrow indicates that the second event follows the first. The following conditions prevailed during testing: Subjects in the intervene group received presentations of X for each lever press they made. The presentation of the test stimulus in the last three conditions was yoked to that of the rats in the intervene condition. A = flashing light; X = tone; F = food (a sucrose solution); LP = lever press; B = click train.

Method

Subjects

Thirty-two experimentally naïve female Long–Evans rats (*Rattus norvegicus*) obtained from Harlan Laboratories (Indianapolis, IN) served as subjects. Subjects were pair-housed in translucent plastic tubs with a substrate of wood shavings in a vivarium maintained on a 12-hr dark/12-hr light cycle. All experimental manipulations were conducted during the dark portion of the cycle. A progressive food restriction schedule was imposed over the week prior to the beginning of the experiment, until each rat received 15 g of food each day. All animals were handled daily for 30 s during the week prior to the initiation of the study. Subjects were randomly assigned to one of four groups ($n = 8$): intervene, observe 1, exogenous cue, and observe 2.

Apparatus

Each of eight experimental chambers measuring 30 × 25 × 20 cm (length × width × height) was housed in a separate sound- and light-attenuating environmental isolation chest (Med Associates, Georgia, VT). The side walls and ceiling of the chamber were constructed of clear Plexiglas. The front and rear walls were constructed of aluminum panels. The floor was constructed of stainless-steel rods measuring 0.5 cm in diameter, spaced 1.5 cm apart center-to-center. The enclosure was dimly illuminated by a 28-V, 100-mA shielded incandescent house light mounted on the top of the rear wall of the conditioning chamber, 2 cm below the ceiling. All experimental procedures were conducted with the house light on, except where otherwise noted. A diffuse light was located 13 cm above the floor, 1 cm below the house light. A flashing light (0.25 s on/0.25 s off) could be presented by flashing the diffuse light. The house light was turned off during the duration of the flashing light presentation. Three speakers on the outside walls of the chamber could deliver a high-frequency tone (3000 Hz) 8 dB(A-Scale) above background, a white noise stimulus 8 dB above background, and a click train (6/s) 8 dB above background. (The noise was not used in Experiment 1.) Each chamber was equipped with a dipper that could deliver sucrose solution (20%). When in the raised position, a small well (0.05 cc) at the end of the dipper arm protruded up into the feeding niche. Sucrose solution served as Stimulus F. An infrared photo detector was positioned across the entrance to the feeding niche. When a rat placed its nose into the feeding niche to lick the sucrose solution, the photo beam was disrupted. The computer measured the disruption of the photo beam, which we refer to as a nose poke. A retractable lever was located 3 cm to the left of the drinking niche, 6 cm above the chamber floor, and could be extended 2 cm into the chamber. Ventilation fans in each enclosure and a white-noise generator on a shelf outside of the enclosures provided a constant 74-dB background noise.

Procedure

Food hopper training. The levers were retracted during all phases of the experiment except during testing. On Day 1 of the experiment, the rats were trained to eat the sucrose out of the food hopper. This was done in a single 60-min session during which a dipper filled with sucrose solution was raised into the food hopper

once every 5–35 s. During each presentation sucrose was available for 10 s, after which the dipper was lowered into the food trough containing sucrose.

Phase 1: A→X presentations. On Days 2–5, all subjects received six daily A→X trials during which Stimulus A was presented for 10 s followed by Stimulus X, which was also presented for 10 s. The flashing light served as Stimulus A, and the tone served as Stimulus X. These trials occurred every 3–7 min during each 35-min session.

Phase 2: A→F presentations. On Days 6 and 7, all subjects received 12 A→F trials in each 65-min session. A trial consisted of Stimulus A presented for 10 s, followed by the delivery of Stimulus F (sucrose) for 10 s. Trials occurred every 2–6 min. To measure acquisition of the A→F association, we recorded the amount of time the rat spent inspecting the feeder (nose poking) during a 30-s period prior to the onset of Stimulus A and in the 10 s during which Stimulus A was present. These measures allowed us to compute a discrimination ratio, which was calculated as the number of nose pokes into the feeder during Stimulus A divided by the sum of nose poking during Stimulus A and during the 30-s period (divided by 3 to produce an equal interval of baseline responding) prior to the onset of Stimulus A. A discrimination ratio above 0.5 indicates that the rats expected F more during Stimulus A than in the baseline interval prior to the onset of A.

Test. Testing was conducted on Day 8 of the experiment. The levers, which had not been present in the conditioning chamber before, were extended into the chambers for the first time during the test session. During the 60-min test session, subjects were tested on Stimulus X in the following manner. Rats in the intervene group received a 10-s presentation of X each time they pressed the lever, with the exception that lever presses that occurred during the presentation of X did not have any consequence (i.e., were not effective). Each rat in the intervene group served as a master rat to which three rats in adjacent chambers—one for each of the observe 1, exogenous cue, and observe 2 groups—was yoked. The subjects in the yoked groups received their test treatments whenever their master rat pressed the lever. Rats in the observe 1 group received a 10-s presentation of X for each effective LP by the master rat in the intervene group. Rats in the exogenous cue group received a 10-s presentation of B (the click) followed immediately by a 10-s presentation of X (termination of B simultaneous with the onset of X) for each effective LP by the master rat in the intervene group. Rats in the observe 2 group received a 10-s period with no programmed stimuli followed by a 10-s presentation of X for each effective LP by the master rat in the intervene group. Observe 2 group's treatment attempted to control for any effects produced by the different temporal intervals between LPs by the master rat and the onset of X in the exogenous cue and observe 2 groups. We planned to pool the data across the observe 1 and observe 2 groups, barring any group differences in mean nose pokes. LPs in the observe 1, exogenous cue, and observe 2 groups had no consequence (that is, the rats in these groups could press the lever, but pressing the lever did not cause the presentation of any stimuli. The levers were made available in these groups to equate the stimulus context with that experienced by the rats in the intervene group). During each test trial, we recorded the number of nose pokes during the 10-s presentation of X. No sucrose presentations were given during the testing phase.

Results and Discussion

Phase 2: A→F presentations. Prior to the factorial analysis, an outlier analysis was conducted to remove from the experiment any subjects with a discrimination ratio two standard deviations from their respective group means on the final day of acquisition. Data from 3 subjects below their group means were removed (because low discrimination scores indicate failure to learn the A→F relationship), resulting in final *n*s = 8 for the intervene and observe 2 groups, $n = 7$ for the exogenous cue group, and $n = 6$ for the observe 1 group. All groups acquired the discrimination. An analysis of variance (ANOVA) conducted on discrimination ratios from the second acquisition session revealed no main effect of group, $F(3, 25) = 2.03$, $MSE = 0.04$, $p > .05$.

Testing. No difference was found in the mean number of nose pokes between the observe 1 and observe 2 groups, $t(12) < 1$; therefore these data were pooled to create the observe-pooled group ($n = 14$). Subjects in the intervene group made significantly fewer nose pokes ($M = 3.79$, $SEM = 1.50$) than did subjects in the observe-pooled ($M = 18.77$, $SEM = 3.81$) and exogenous cue ($M = 17.53$, $SEM = 5.69$) groups, which did not differ from each other (see Figure 2). A one-way ANOVA conducted on mean nose pokes during X revealed a main effect of group, $F(2, 26) = 3.90$, $MSE = 158.61$, $p < .05$. Planned comparisons using the error term from the one-way ANOVA were conducted to determine the source of the main effect. Responding in the intervene group was lower than responding in the observe-pooled group, $F(1, 26) = 7.21$, $MSE = 158.61$, $p < .05$, replicating the effect of a lever press intervention reported by Blaisdell et al. (2006). Interestingly, the intervene group also responded less than did the exogenous cue group, $F(1, 26) = 4.38$, $MSE = 158.61$, $p < .05$, suggesting that when Stimulus B preceded Stimulus X at test, it did not disrupt nose-poke responding to X. Mean nose pokes in the exogenous cue and observe-pooled groups did not differ, $F(1, 26) < 1$, $MSE = 158.61$. Finally, mean LPs in the intervene ($M = 29.38$, $SEM = 16.45$), observe-pooled ($M = 18.93$, $SEM = 4.13$), and exogenous cue ($M = 25.29$, $SEM = 5.58$) groups did not differ, $F(2, 26) < 1$, $MSE = 752.39$.

These results replicate the effect reported by Blaisdell et al. (2006; Experiments 1 and 2a) that an LP preceding a stimulus at test attenuates nose poking for food (F) if the stimulus was trained

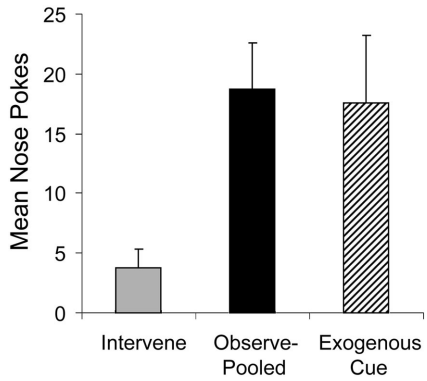

Figure 2. Mean nose pokes elicited during X for the intervene, observe-pooled, and exogenous cue groups in Experiment 1. Error bars represent standard errors of the mean.

as an effect of a common cause, A, which also causes F. This result is consistent with causal-model theory (Waldmann et al., 2008; Waldmann & Hagmayer, 2005), which predicts that the LP should be viewed as a potent alternative cause of Effect X, which in the learning phase was caused by A. Apparently very few trials are needed to establish an action as a potent alternative cause leading to discounting. This is consistent with the assumption that actions are regarded as independent of the system they target. The novel finding of Experiment 1 is that the same number of test trials did not suffice to establish an exogenous stimulus, B (a click train), as causal. Hence, no discounting was observed for this cue. The lack of an effect of B on responding to X at test also provides evidence against retroactive interference as a source of the attenuated responding to X produced by the LP in the intervene group.

Experiment 2

Although the results of Experiment 1 are compelling, one caveat needs to be addressed. In Experiment 1, we equated the durations of A (light) and B (click train), which were both 10 s, because we thought it likely that for B to effectively interfere with A, B should be similar to A. If, however, we consider the notion that the LP→X (tone) relationship at test may have retroactively interfered with retrieval of the A→X association, we must recognize that the durations of the LP and of B are quite different. LPs are short and of variable duration, typically ranging from 0.40 s to 2.0 s, while A was always fixed at a relatively longer 10 s. Different durations may also be viewed as a potential confound when testing whether rats treat actions different from observed events. Although this possibility seems somewhat far fetched, maybe briefly presented events are more easily established as causal than are events with longer duration.

Thus, Experiment 2 served as a replication of Experiment 1, with each group trained with the A→X and A→F (food) pairings, except that we equated the temporal parameters of the external Cause B, which was introduced at test, with the LPs at test (see Table 2). This was accomplished by yoking the onset and termination of B at test in the exogenous cue group to the onset and termination of the LP emitted by the master rat in the intervene group.

We also included an unpaired control group in Experiment 2 to establish the necessity of the A→F pairings on the expectation of F during test trials of X in the observe group. There are many examples in the literature on Pavlovian conditioning of an increased response to a stimulus not because it was paired with a motivating outcome, but simply because the motivating outcome had been repeatedly presented (i.e., sensitization). For example, if dogs receive presentations of a bell paired with food, the bell might acquire the ability to elicit salivation when sounded on its own. The typical explanation is that dogs learned that the bell signals food and thus expect food when they hear the bell. An alternative explanation is that the mere repeated presentation of food causes dogs to become excited and that when the bell is played at test, dogs are salivating merely because they are in a higher state of arousal. To rule out this alternative sensitization account, Pavlovian psychologists have typically included a group of dogs (in this example) for which the bell and the food are presented but in an unpaired fashion—that is, they are not presented one followed by the other but are separated from each other by many minutes. If the dogs in this unpaired control group do not develop the salivation response to the bell, then this suggests that the experimental group (which got the bell→food pairings) developed the response because the dogs learned the bell–food association and thus anticipated the food when they heard the bell.

We wished to demonstrate that, in our procedure, the high rate of nose poking in the observe group was due to the fact that the rats expected F during X and not simply because they had received presentations of the food during Phase 2. That is, we wished to establish through a control group that nose poking in the observe group was excitatory in nature. In an additional attempt to evaluate excitatory control by X in the test phase, we collected nose poke responses during the 10 s prior to the onset of X at test, as well as during X. This allowed us to compute a discrimination ratio (as we did during Phase 2), which can be used to determine if the rate of nose poking during X at test is driven by an expectation of F. If it is, then responding during X should be higher than during the pre-X interval, which establishes the baseline rate of nose poking. Furthermore, the collection of pre-X nose pokes also allowed us to compare baseline rates of nose poking across groups, which should be uniformly low.

Table 2
Design of Experiment 2

Group	Phase 1	Phase 2	Test
Intervene	A→X	A→F	LP→X
Observe	A→X	A→F	X
Exogenous cue	A→X	A→F	B→X
Unpaired	A→X	A / F	X

Note. An arrow indicates that the second event follows the first, and events on either side of a slash mark were presented at separate times within the same session. A = flashing light; X = noise; F = food (a sucrose solution); LP = lever press; B = tone.

Method

Subjects and Apparatus

Forty-eight rats of the same type and maintenance as in Experiment 1 were used. Subjects were randomly assigned to one of four groups: intervene, observe, exogenous cue, and unpaired, *n*s = 12 per group. The apparatus was the same as that used in Experiment 1, except that the noise served as X and the tone served as B in this experiment.

Procedure

Food hopper training. The levers were retracted during all phases except testing. On Day 1, sucrose was presented every 5–35 s in the 60-min session.

Phase 1: A→X presentations. On Days 2–5, all subjects received six daily A→X presentations during which Stimulus A was presented for 10 s followed by the presentation of Stimulus X for 10 s. Trials were delivered every 3–7 min during each 35-min session.

Phase 2: A→F presentations. On Days 6–7 treatment was as in Experiment 1 for subjects in the intervene, observe, and exog-

enous cue groups, with 12 trials in each 65-min session. On each trial, Stimulus A was presented for 10 s followed by the presentation of Stimulus F for 10 s. Subjects in the unpaired group received 12 presentations of Stimulus A and 12 presentations of Stimulus F, each for 10 s in an explicitly unpaired fashion. That is, presentations of A and of F were separated by a mean of 1–3 min. As in Experiment 1, we collected nose pokes during A and during the 30 s prior to A so as to calculate discrimination ratios for each subject.

Test. Levers were inserted into the chambers prior to the test phase. Subjects received the same treatment on Day 8 as in Experiment 1, except for the following. Each subject in the exogenous cue group was yoked to a master rat in the intervene group so that each LP by the master rat initiated Stimulus B followed by X. On these test trials, B onset at the same time that the master rat pressed the lever and B terminated at the same time that the master rat stopped pressing the lever. Thus, the onset and duration of each presentation of B for a yoked subject in the exogenous cue group precisely matched (within the temporal resolution of the computer) the onset and duration of an LP for a master subject in the intervene group. Nose pokes were recorded for each subject (yoked and master) during each 10-s interval prior to each effective LP by the master rat. Nose pokes were also recorded during each 10-s presentation of X at test for all subjects. These measures allowed us to calculate a discrimination ratio to measure the excitatory strength (i.e., the degree to which the rat expected F) of responding to X at test. We also recorded the total number of LPs by each rat in the 60-min test session. As is conventionally done (e.g., Blaisdell, Denniston, & Miller, 1998), prestimulus and stimulus scores from test trials were transformed to log (base 10) scores to better fit the assumption of parametric statistics concerning normal distributions of scores within groups.

Results and Discussion

A→F presentations. An outlier analysis conducted on mean discrimination ratios calculated from responses during training trials on the 2nd day of acquisition for subjects in the intervene, observe, and exogenous cue groups resulted in the removal of 2 subjects—1 from the observe group and 1 from the exogenous cue group—whose scores were two standard deviations below their respective group means. A separate outlier analysis was conducted in the same manner for the unpaired group, but no outliers were identified.

Discrimination scores for the remaining subjects in the intervene, observe, and exogenous cue groups increased across sessions, indicating learning of the A→F relationship. As expected, however, subjects in the unpaired group failed to show changes in discrimination ratios across sessions. These observations were supported by a one-way repeated-measures ANOVA conducted on discrimination ratios with group as a between-subjects variable and session as a repeated measure, which found a main effect of session, $F(1, 42) = 48.42$, $MSE = 0.011$, $p < .001$, and of group, $F(3, 42) = 7.87$, $MSE = 0.011$, $p < .001$, and a Session × Group interaction, $F(3, 42) = 5.17$, $MSE = 0.011$, $p < .01$. To find the source of the interaction, we conducted planned comparisons using the error term from the ANOVA. These analyses revealed a change in the discrimination ratio across sessions for the intervene, observe, and exogenous cue groups, Fs(1, 42) = 7.26, MSEs = 0.011, ps < .05. As expected, discrimination ratios did not change across sessions for the unpaired group, $F(1, 42) < 1$, $MSE = 0.011$.

Testing. It is not possible to log-transform zero scores, and thus the data from 3 subjects were eliminated from testing. The remaining subjects were included in the following statistical analyses, with $n = 11$ in each of the intervene, exogenous cue, and unpaired groups, and $n = 10$ in the observe group. To determine whether there were any group differences in baseline responding, we conducted a one-way ANOVA on the mean number of prestimulus nose-poke responses, which revealed no group differences, $F(3, 39) = 1.52$, $MSE = 0.33$, $p > .22$.

Figure 3 shows mean discrimination ratios for nose poking on test trials with Stimulus X for each group. Responding in the observe and exogenous cue groups was higher than in the intervene group, replicating the results of Experiment 1. Furthermore, the observe group responded more than did the unpaired group, indicating the necessity of the A→F pairings on excitatory food-anticipatory responding to X at test in the observe group. These findings were supported by statistical analyses. A one-way ANOVA conducted on mean discrimination ratios during X at test revealed a main effect of group, $F(3, 39) = 3.93$, $MSE = 0.001$, $p < .02$. Planned comparisons using the error term from the one-way ANOVA were conducted to determine the source of the main effect. Discrimination ratios in the observe group were higher than in the intervene and unpaired groups, Fs(1, 39) ≥ 7.28, $MSE = 0.01$, ps < .01. Subjects in the exogenous cue group also showed marginally higher discrimination ratios than did subjects in the unpaired group, $F(1, 39) = 3.85$, $MSE = 0.01$, $p = .057$. Mean discrimination ratios in the exogenous cue and observe groups did not differ, $F(1, 39) = 1.05$, $MSE = 0.01$, $p > .30$. Likewise, discrimination ratios in the intervene and unpaired groups did not differ, $F < 1$.

As a second measure of excitatory (i.e., above base rate) performance during testing, t tests for single means were used to assess group performance against a discrimination ratio of 0.5—a value that reflects no change in rates of responding during the test

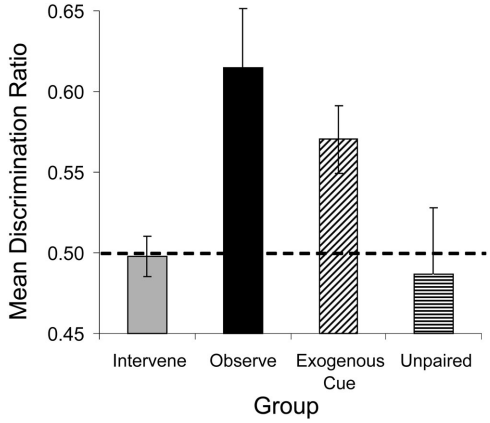

Figure 3. Mean discrimination ratios calculated from log-transformed scores at test for the intervene, observe, exogenous cue, and unpaired groups in Experiment 2. Error bars represent standard errors of the mean. Dashed line indicates no change in rate of responding to the test stimulus relative to base rate levels of responding.

stimulus compared with baseline rates assessed during the pre-stimulus interval. Discrimination scores in the observe and exogenous cue groups were significantly greater than 0.5, $t(9) = 3.01$, $p < .02$, and $t(10) = 3.12$, $p < .01$, respectively. Discrimination scores in the intervene and unpaired groups did not differ from 0.5, $ts(10) < .32$. As in Experiment 1, mean LPs in the intervene ($M = 23.18$, $SEM = 6.04$), observe ($M = 15.20$, $SEM = 2.83$), exogenous cue ($M = 21.55$, $SEM = 5.58$), and unpaired ($M = 20.00$, $SEM = 3.31$) groups did not differ, $F(3, 39) < 1$, $MSE = 239.85$.

Experiment 2 replicated both main findings of Experiment 1: (a) an LP served as an effective alternative cause and attenuated nose-poke responding to X at test and (b) an exogenous Stimulus B did not affect nose-poke responding to X and thus failed to be established as an independent cause leading to discounting. One difference from Experiment 1 was that the durations of B at test were matched to the durations of the LPs. The lack of an attenuation of excitatory responding to X at test despite the similarity between the durations of B and lever pressing provides further evidence for the special status of actions and against a retroactive interference account of the effect of an LP intervention.

Another novel result of Experiment 2 was that the calculation of discrimination ratios at test allowed us to directly show the excitatory nature of nose-poke responding to X at test. Although baseline (pre-X) levels of responding were similar in all groups, only the observe and exogenous cue groups showed an elevation in nose-poke responding during X. The inclusion of an unpaired control group further established the excitatory nature of responding to X at test in the observe group. Although a similar group was included in Blaisdell et al. (2006; Experiment 2b), our replication strengthens confidence in this interpretation. The results of Experiments 1 and 2 lend further support to the hypothesis that actions are a special kind of event for rats.

Experiment 3

One important further difference between retroactive interference theories and the causal-model view concerns possible transfer effects. Causal-model theory predicts that discounting due to an intervention should be restricted to the moment of action and not transfer to later tests in which subjects are presented with different causal test situations. Thus, a causal reasoner should be capable of switching back and forth between correct predictions on the basis of actions or external events without being influenced by previous predictions. In contrast, retroactive interference has been shown to last up to 3 days after the second, interfering association is acquired (Escobar et al., 2001; Experiment 2b). Thus, if attenuation in nose poking observed in the intervene group of Experiments 1 and 2 was due to retroactive interference produced by a strong LP→X (tone) association acquired on the 1st test day, then we should continue to observe attenuated nose pokes to X on a subsequent session during which X is presented alone (i.e., in the absence of a preceding LP).

The following experiment tests these predictions by presenting both observation and intervention test conditions to each subject. All subjects in this experiment learned about a common-cause model during training. That is, all rats learned that stimulus A (light) is a common cause of X and F (food). Additionally, a separate cue, Y (noise), was established as an alternative cause of F in the learning phase (as in Blaisdell et al., 2006, Experiment 1). Following training, rats were tested on either X or Y. Half the rats that were tested on X received the intervene test condition on the 1st test day followed by the observe test treatment on the 2nd test day. The remaining rats that were tested on X received the same two test treatments in the reverse order. If the LP intervention on X attenuates nose poking during X due to retroactive interference produced by the acquisition of an LP→X association on Test Day 1, then we should continue to observe attenuated nose poking when subjects are tested on X in the absence of lever pressing on Test Day 2. By contrast, causal-model theory predicts that the LP intervention test on Test Day 1 will not result in attenuated nose poking during presentations of X on Test Day 2. This is because the rat should be able to discriminate whether presentations of X were caused by its own action (as on the intervention test) or not by its own action (as on the observation test). Only when presentations of X are attributed to the rat's own action does causal-model theory predict attenuated responding to X. In the condition in which the rats received the observe test on Test Day 1 and the intervene test on Test Day 2, the retroactive interference account and causal-model theory both predict more nose pokes to the outcome on Test Day 1 than on Test Day 2. The design of Experiment 3 is shown in Table 3.

All subjects also received training on a direct-cause model as in Blaisdell et al. (2006; Experiment 1) in which Stimulus Y (e.g., a noise) is paired directly with F. Subjects in the direct-cause test conditions were tested on Stimulus Y instead of Stimulus X. Half the subjects in this condition received Y in the intervene test on Test Day 1 and the observe test on Test Day 2, while the remainder of the subjects received the observe test first followed by the intervene test. Blaisdell et al. showed that by pairing Y directly with F, subjects treated Y as a direct cause of F (in contrast, X predicts F only through association with A). This led rats to expect F when subsequently tested on Y, regardless of whether Y had been produced by an LP intervention or was merely observed. This result is consistent with causal-model theory, which predicts that interventions on a direct (or indirect) cause of an effect (e.g., food) should not lead to discounting. The direct-cause groups were included in Experiment 3 as a further test that rats do not generally decrease their expectation of food after LPs. Moreover, the conditions served as a control for a response-competition account of the attenuating effects on nose pokes after an intervention. If the lower rate of nose poking following an LP intervention is due to response competition between lever pressing and nose poking, then we would expect to observe a low rate of nose poking

Table 3
Design of Experiment 3

Group	Phase 1	Phase 2	Test 1	Test 2
Common-intervene	A→X	A→F / Y:F	LP→X	X
Common-observe	A→X	A→F / Y:F	X	LP→X
Direct-intervene	A→X	A→F / Y:F	LP→Y	Y
Direct-observe	A→X	A→F / Y:F	Y	LP→Y

Note. An arrow indicates that the second event follows the first, events on either side of a slash mark were presented at separate times within the same session, and a colon indicates that both events occurred simultaneously. A = flashing light; X and Y = tone and noise, respectively, counterbalanced; F = food (a sucrose solution); LP = lever press.

following an intervention on both X and Y. If the effect of an LP intervention operates through causal discounting, then we expect a lower rate of nose poking when the LP causes X but not when it causes Y.

Method

Subjects and Apparatus

Thirty-two rats of the same type and maintenance as in Experiment 1 were used. All subjects received both common-cause and direct-cause training, but only one of the two models was investigated in the test phases. Hence, subjects were randomly assigned to one of four test conditions: common-cause-intervene-first, common-cause-observe-first, direct-cause-intervene-first, and direct-cause-observe-first, *n*s = 8 per group. The apparatus was the same as that used in Experiment 1. The flashing light served as Stimulus A. Tone and noise were counterbalanced within groups in their roles as Stimulus X and Y, respectively.

Procedure

Food hopper training. As in Experiment 1, the levers were retracted during all phases except testing. On Day 1, sucrose was made available to the rat by raising the dipper arm containing sucrose every 5–35 s in a 60-min session.

Phase 1: A→X presentations. As in Experiment 1, on Days 2–5, all subjects received six daily A→X presentations during which Stimulus A was presented for 10 s followed by the presentation of Stimulus X for 10 s. Trials were delivered every 3–7 min during each 35-min session.

Phase 2: A→F and Y:F presentations. On Days 6 and 7, all subjects received 12 A→F pairings as in Experiment 1, interspersed among 12 Y–F simultaneous pairings in each 65-min session. On Y–F trials, Y and F were presented simultaneously for 10 s, such that the onset and termination of both stimuli coincided. We decided to simultaneously present Y and F because this allows us to better compare the Y stimulus with the X stimulus, which is involved in the common-cause model, in which X and F should also be expected to occur simultaneously. Trials occurred every 1–3 min.

Test. Testing was conducted on Days 8 and 9 as in Experiment 1 except for the following: Levers were extended into the chambers only during the intervention tests and were absent during the observation tests. This was done to prevent extinction of lever-pressing for the groups that observed a stimulus on Day 8 (the 1st test day) but were tested on the intervention test condition on Day 9 (the 2nd test day). Half of the subjects in the common-cause group were tested in the intervention test as in Experiment 1, such that each LP resulted in the 10-s presentation of X. The remaining subjects received presentations of X yoked to LPs of a master rat in the same group that received an intervention test. Half the subjects in the direct-cause group received a 10-s presentation of Y for each effective LP. The remaining subjects in the direct-cause group received presentations of Y yoked to the LPs of a master rat in the same group receiving an intervention test. On Day 9, subjects within each group received the alternate test treatment from what they had received on Day 8. Thus, subjects tested on the intervention test on Day 8 received the observation test on Day 9, and vice versa. Nose pokes during each stimulus presentation were recorded.

Results and Discussion

A→F / Y treatment. Separate outlier analyses conducted on mean nose pokes during A and Y revealed two outliers, one above and one below their respective group mean. The outliers were excluded from all statistical analyses, resulting in a final $n = 7$ for the common-cause-intervene-first and common-cause-observe-first groups and an $n = 8$ for the direct-cause-intervene-first and direct-cause-observe-first groups. A one-way ANOVA conducted on mean discrimination ratios from Day 2 of acquisition, with group as a between-subjects variable and cue type (A or Y) as a within-subject variable, revealed no main effects or interaction, Fs(3, 26) < 1.

Testing. Figure 4 shows mean nose pokes during X or Y. On both days subjects that intervened on X nose-poked less than did subjects that observed X, while no difference in mean nose pokes

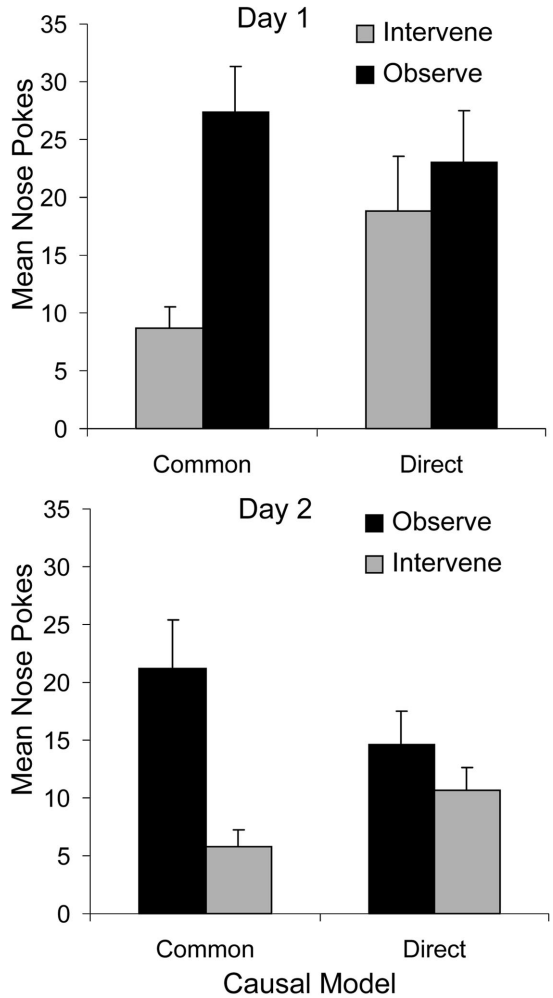

Figure 4. Mean nose pokes elicited during X (common-cause conditions) and Y (direct-cause conditions) at test following a lever press (intervene) or no lever press (observe) on Test Days 1 (top panel) and 2 (bottom panel) in Experiment 3. Error bars represent standard errors of the mean.

were found between subjects that intervened on or observed Y. More importantly, subjects that showed low rates of nose poking when their LPs produced X on the 1st test day showed higher rates of nose poking when they observed X on the 2nd test day. These interpretations are supported by statistical analyses. A three-way mixed ANOVA with causal model (common-cause and direct-cause) and test order (intervene or observe condition tested first) as between-groups variables and test day (1st and 2nd) as a within-subject variable conducted on mean nose pokes during the test stimulus revealed a main effect of day, $F(1, 26) = 5.86$, $MSE = 104.27$, $p < .03$, a two-way interaction between test order and day, $F(1, 26) = 15.82$, $MSE = 104.27$, $p < .001$, and a three-way interaction between causal model, test order, and day, $F(1, 26) = 5.96$, $MSE = 104.27$, $p < .03$. Planned comparisons were conducted to isolate the source of the three-way interaction. Rats that intervened on X (from the common-cause model) nose-poked less than did rats that observed X on both Test Day 1, $F(1, 26) = 9.59$, $MSE = 104.27$, $p < .005$, and Test Day 2, $F(1, 26) = 12.81$, $MSE = 104.27$, $p < .005$, while rats tested on Y (from the direct-cause model) did not show different rates of nose poking in either test condition on either test day, Fs $(1, 26) < 1$, $MSE = 104.27$. An additional two-way mixed ANOVA with causal model (common-cause and direct-cause) as a between-groups variable and test condition (intervene and observe) as a within-subject variable conducted on mean nose pokes pooled across both test days revealed a main effect of test condition, $F(1, 28) = 13.59$, $MSE = 121.42$, $p < .001$, and a Causal Model × Test Condition interaction, $F(1, 28) = 5.12$, $MSE = 121.42$, $p < .05$. Planned comparisons were conducted to isolate the source of the two-way interaction. In the common-cause condition, rats made fewer nose pokes in the intervene test condition than in the observe test condition, $F(1, 28) = 16.59$, $MSE = 121.42$, $p < .001$, while there was no difference in mean nose pokes between these test conditions for rats in the direct-cause condition, $F(1, 28) = 1.08$, $MSE = 121.42$, $p > .05$.

On Day 1 of testing, rats in the intervene test conditions made marginally fewer LPs when tested on X ($M = 22.14$, $SEM = 5.12$) than when tested on Y ($M = 37.63$, $SEM = 10.62$), $t(13) = 1.25$, $p = .07$. Lever pressing did not differ between these groups on Day 2 of testing, $t(13) < 1$. No levers were present for rats in the observe test conditions.

We observed that lever press interventions attenuated nose poking to X but not to Y. This difference provides evidence against response competition as an explanation for the attenuated nose pokes to X following an LP. If lever pressing generally interfered with nose poking, we would have observed depressed responding in both LP conditions. More importantly, using a within-subject test we observed rats to nose-poke at a higher rate when they observed X on the second test session even though they nose-poked at a lower rate when they intervened on X in the first test session. This finding fails to support the hypothesis that the low rates of nose poking on the intervention test are due to the acquisition of an LP→X association that retroactively interfered with the A→X association acquired in Phase 1. If that were the case, then we would also have expected interference by the LP→X association established on Day 1 in the observe test of X on Day 2.

Rather, these results support causal-model theory by showing that the expectation of F during X depended on whether X had been observed or had been produced by an LP in the particular test situation. This was true whether rats received the intervention test condition first and the observation test condition second or the reverse. These data suggest that rats are able to make flexible causal inferences that depend not on the storage and retrieval of stable associations but on the prevailing conditions at the time of test. Inferences drawn from interventions on an event can differ from those drawn from separate observations of the same event—a strong aspect of a causal reasoning process.

Meta-Analysis

A key prediction of causal-model theory is that just a few trials should suffice to establish an action as a cause. This is a consequence of the assumption that actions are typically unconfounded, so that a single learning instance often suffices to generate valid causal inferences. Hence, if rats experience spatiotemporal pairings between their actions and otherwise rare outcomes (e.g., a tone X), it should not take the rat a significant amount of time to reason that it had caused Stimulus X by pressing the lever. Rather, this inference might be readily apparent on the very first instance of lever pressing. In contrast, arbitrary events are more likely to be confounded so that a larger sample size and more observations about surrounding events are necessary to establish a firm causal relation.

Therefore it would be informative to look at the effect of an LP intervention on nose poking during X on the very first test trial. We conducted a meta-analysis of first-trial test data from four experiments to assess the effect of a lever press intervention on X compared with the first observation of X test trials. Although each study contained additional groups for various purposes, they all contained one group of rats that had received the intervene test with X and another group that had received the observe test with X after they had received common-cause training using the procedure of Blaisdell et al. (2006; Experiment 2a) to train the common-cause model X←A→F.

The meta-analysis was performed on the nose-poke scores from the first test trial on Stimulus X from 36 subjects in the intervene test conditions and 34 subjects in the observe test conditions. These subjects were pooled from four separate studies, including Blaisdell et al. (2006; Experiment 2a, $n = 20$), Experiment 1 of this article ($n = 14$), Experiment 2 of this article ($n = 22$), and Experiment 3 of this article (data from only the 1st test day were included; $n = 14$). Overall, subjects in the intervene test condition made fewer nose pokes ($M = 4.31$, $SEM = 1.63$) than did subjects in the observe test condition ($M = 13.15$, $SEM = 3.16$). This difference was reliable, $t(68) = 2.53$, $p < .001$. These data suggest that rats are capable of reasoning from causal interventions on the very first opportunity to intervene on the effect.

This finding also further weakens the retroactive inference account. In previous demonstrations of retroactive interference, the number of B→O pairings ranged from 4 (Escobar et al., 2001) to 10 (Matute & Pineño, 1998). Associative theories suggest that the associations between an LP and X accrue over trials. On the contrary, the predictions generated from a causal-model account would predict that an intervention can immediately sever the connection between A and X (see Pearl, 2000; Spirtes et al., 1993; Woodward, 2003), if lack of confounding is assumed.

General Discussion

The results of our experiments provide further evidence that rats are capable of causal reasoning (see also Blaisdell et al., 2006; Waldmann et al., 2008). The findings indicate that rats differentiate between causal and noncausal covariations and adapt their observational and interventional expectations accordingly. These results contradict the view that only humans among all animals are capable of causal reasoning (Gopnik & Schulz, 2004; Penn & Povinelli, 2007; Povinelli, 2000). The present findings demonstrate that rats understand that both noncausal and causal relations can be used to make observational predictions, but that only direct or indirect causal effects can be affected by interventions. The present findings also show that rats can flexibly switch between observational and interventional predictions (Experiment 3) and that they draw inferences from the causal properties of their actions in the very first intervention test trial (Meta-Analysis).

An important new focus of the present set of studies was the special status of actions compared with other events. In the framework of causal-model theory, any causal event that is introduced in the test phase can play the role of an alternative cause leading to discounting of a previous cause (Waldmann et al., 2008; Woodward, 2003). Full discounting is expected, however, only when the new event is established as a deterministic and independent cause. Thus, if rats had received extensive prior training showing them that Event B perfectly covaries with X and furthermore is independent of the alternative causes of X, then B would lead to discounting similar to the interventions in X. Confronted with very few trials at test and the potential confounding with other factors, however, we can readily explain why arbitrary events such as the click B will not be established as a strong alternative cause during the brief test phase. Given that under these circumstances B and X are not causally related, no discounting is to be expected. In contrast, self-generated actions are more likely to be considered free and independent (Haggard, 2005; Wegner, 2002), at least by humans, which implies that they are typically not confounded with other causes (see Blaisdell, 2008, for an extended discussion and Killeen, 1981, and Killeen & Smith, 1984, for a related demonstration in pigeons). Consequently, a single trial or very few trials will suffice to create a strong causal impression.

Note that this analysis is not inconsistent with the assumption that alternative causes, regardless of whether they represent interventions or arbitrary events, are often viewed as independent (Cheng, 1997; Hagmayer & Waldmann, 2007). Once an event has been established as a strong cause, independence might be indeed a plausible default assumption. Our point is that the independence assumption also plays a role during the learning phase, in which an event is established as a potential cause. Whenever we assume independence, as with our own actions, fewer trials than with potentially confounded events are needed to establish a causal relation. The lack of discounting we see when arbitrary events were paired with the target effect in only a few test trials can be explained by the assumption that such brief exposure does not suffice to establish a strong causal relation for these events.

We know that humans view themselves as capable of free, independent interventions (e.g., Wegner, 2002). The present results indicate that actions are special for rats as well, although it is, of course, unclear whether they are in any sense aware of the special status of actions. The special status of actions was particularly prominent in Experiment 3, in which the same subject had an opportunity to intervene on X or to observe X on separate occasions. It was only when X had been produced by their lever press intervention that the rats had discounted the previously established Cause A. Thus, rats treated their LPs as having the causal power to fix the variable of X to the *on* state, whereas in the absence of an LP they appeared to have attributed the change in the value of X from *off* to *on* to the usual Cause A.

Intentional actions are so important that specialized neural systems exist to monitor actions as they are executed (Sirigu et al., 2004). These systems take advantage of the many unique sensory feedback cues—kinesthetic, proprioceptive, haptic, visual, etc.—that result from planned action and allow the individual to correct errors that may occur during the execution of the action. These sensory feedback cues are also responsible for the sense of an "I" (the agent) that controls events in the outside world (Haggard, 2005). In fact, the temporary inactivation by transcranial magnetic stimulation of particular brain areas involved in monitoring self-generated actions, such as the presupplementary motor area, renders subjects incapable of distinguishing between effects of their own actions and effects of exogenous causes (Haggard and Clark, 2003; Haggard, Clark, & Kalogeras, 2002). These internal sensory markers of intentional behavior are likely critical for the ability to reason from causal interventions. Reasoning from an intervention requires the subject to distinguish the effects of its actions from other events; thus, without the feedback cues and self-monitoring necessary to make this distinction, the agent would be incapable of accurate reasoning from its own interventions. These circuits may have developed to enable accurate self-generated movement, but they may have also developed to monitor the effects of actions on the world and thus determine their causal power. The internal sensory feedback signals may also be responsible for an increase of salience of actions, which may mediate faster learning of instrumental action–effect relations (Blaisdell, 2008).

The possibility of higher salience of actions raises the question of how much knowledge humans and especially nonhumans have about their actions. Actions are special because they are typically independent (i.e., free) and often deterministic. Without these features it would not be justified to attribute causality to actions as quickly as humans as well as rats do, regardless of the greater salience of actions. Humans are often aware of these features, at least implicitly (Wegner, 2002). Although humans may not know about the underlying statistical relations, a sense of freedom may accompany many self-initiated actions. This assumption also underlies the reasoning of researchers who plan experiments while assuming that their manipulations are independent of the system under study (Pearl, 2000). We do not know whether rats have a comparable sense of freedom. Their bias to trust instrumental contingencies more than observational contingencies may be a successful heuristic adaptation to how events are typically related in the world. An interesting test case for whether the underlying causal-model representations are sensitive to the statistical boundary conditions of the difference between actions and other events might be to explore the flexibility of causal reasoning with actions that lack determinism or independence. In these cases, less discounting should be observed, according to causal-model theory. Thus far we do not know whether humans, let alone rats, are sensitive to these boundary conditions.

One interesting finding, already observed in Blaisdell et al. (2006), is the apparent dissociation between nose poking and LPs in the direct-cause relation. One would expect that the causal expectancies of food during the direct-cause stimulus, indicated by high rates of nose poking to check for food, should eventually lead to increased rates of LPs as a result of the high contingency between the LP and the direct-cause stimulus and between the direct-cause stimulus and food (i.e., conditioned reinforcement). In other words, if Y is a direct cause of F, and lever pressing produces Y, then rats should be motivated to lever-press often to produce Y and thus its effect F. It is well known, however, that this type of conditioned reinforcement is parameter-dependent (e.g., number of trials, salience, etc.) and not easy to obtain (Mackintosh, 1974; Winterbauer, 2006). Moreover, it should be noted that we presented levers only in the test phase in which no reinforcement (i.e., food) was presented, so that a tendency to act might have been counteracted by extinction processes. Nevertheless, the results raise the interesting question of whether we need to differentiate between causal expectancies and action plans that are based on these expectancies. The possible dissociation between expectancies and action plans is reminiscent of similar dissociations observed in infants. For example, infants often show in habituation, which taps into their expectations, that they are sensitive to the presence of hidden objects (i.e., object permanence), even when their search behavior seems to contradict this conclusion (Baillargeon, Spelke, & Wasserman, 1985).

A further goal of the present studies was to provide tests between causal and associative theories. The finding that rats expect F (food) when at test they observe X (tone) but not when they generate X by means of an intervention is inconsistent with the predictions of acquisition-based associative learning theories, regardless of how the rapid learning of action–effect contingencies is explained (see above). These learning theories would predict attenuated nose poking to X only after an LP if the rat had previously learned that F never accompanies X when generated by an LP. Because the rats in our studies never received such training in the learning phase, they should not have made the distinction at test between observing X and intervening on X with an LP. Other theories, including response-competition theories, are also refuted by the data. No general effect of LPs on nose poking was found in any experiment; only when subjects lever-pressed to produce an effect (X) of an alternative cause (A) was discounting observed (see also Blaisdell et al., 2006). Lastly, retrieval-oriented theories (e.g., Matute & Pineño, 1998) were also not able to accommodate the current data or Blaisdell et al.'s (2006) findings. In particular, we did not find evidence that the occurrence of an exogenous Event B prior to X at test (B→X) retroactively interferes with retrieval of the A→X association. This failure was found both when exogenous cue B's temporal properties matched that of the common Cause A (Experiment 1) and when they matched that of the LP (Experiment 2). Experiment 3 established that prior LP→X experiences in other trials did not interfere with observations of X in later test phases. This result is also inconsistent with a retroactive interference account, which would predict a lasting interference effect by the LP→X association on subsequent observation test trials with X.

Finally, a meta-analysis of first-trial performance, in which subjects observed the co-occurrence of LP and Stimulus X for the first time, precluding any type of prior learning of an association between lever pressing and X, revealed that an LP intervention strongly attenuated nose poking to X on the very first trial. These results clearly demonstrate the special status of actions in causal inferences. When actions are involved, one trial suffices to establish a causal impression, which then leads to discounting.

So far the discussion has focused on rats' inferences about their actions and other events in the test phase (i.e., causal reasoning). The experiments also have a learning component that needs to be discussed. In all studies, second-order conditioning procedures were used to establish common-cause models. Second-order conditioning involves the subject first learning that a relatively neutral stimulus (e.g., a bell) signals a biologically important or significant event (e.g., food or foot shock). The subject then learns that a second relatively neutral stimulus (e.g., a light) signals the first one (the bell). Following these two learning phases, the subject expects food (or a shock) not only when it hears the bell, but also when it sees the light. In fact, our results indicate that rats had learned to expect F (food) when they observed X (tone) at test, even though X had a second-order relationship to F. One question often raised in the context of second-order conditioning is why excitatory relations are formed between the indirectly related events (e.g., X and F) although in the learning phase they were in fact negatively correlated. During learning, A was paired with X in the absence of F, or with F in the absence of X.

One candidate for a normative account of learning and reasoning is causal Bayes nets (see Gopnik et al., 2004; Pearl, 2000; Spirtes et al., 1993). In fact, the idea to compare observational and interventional inferences at test was inspired by causal Bayes net theories, which give a ready computational account of the normative differences between these two types of predictions. Causal Bayes net theories integrate accounts of learning causal structures with formal theories of how these models can be used for reasoning about observations and interventions. The learning theories assume that organisms use information about the full covariation pattern under study to make inductions about the likely generating causal models. Second-order learning of common-cause models or of causal chains presents a problem for this account because the negative correlation between X and F in the learning phase is inconsistent with these two models, which the rats apparently have induced.

Waldmann et al. (2008) have therefore proposed a simpler attention-based account of this learning task. According to this theory, the primary focus during learning is on individual effects that can be caused by alternative causes (cf. Cheng, 1997). We found, based on temporal order and statistical information, that during the observational learning phase rats learn about the relations between Cause A (which is presented temporally prior to its effects) and either effect (X or F). Due to attention limitations, however, they do not focus on both effects at once during this phase. Thus, they learn about two separate causal links without making assumptions about how these two relations are related to each other. This is consistent with findings that show that noticing the negative correlation between indirectly related events in second-order conditioning (e.g., X and F in our procedure) requires many more learning trials than we provided to our rats (Yin et al., 1994).

In the observational test phase, rats are presented with X as a cue. According to Waldmann et al.'s (2008) theory, they then reason link by link from X to A, and since A is part of a further

link, from A to F. Thus, although learning was focused on individual links, the chaining of inferences in the test phase (from X to A to F) effectively leads to inferences consistent with a positive correlation of X and F. The results of the intervention tests can also be modeled within this theory. Viewing interventions as deterministic and independent of the alternative causes of X should lead to full discounting of A. Consequently, the expectation of F, one of the effects of A, should go down. Unlike causal Bayes net theories, this is a more parsimonious, less complex rational account of causal learning, which takes into account potential attentional constraints and predicts biases in more complex scenarios (see Waldmann et al., 2008). Nevertheless, it is a causal (as opposed to associative) theory, which is sensitive to the distinction between cause and effect, noncausal and causal relations, and observations and interventions.

The present studies show that rats are capable of forming causal models relating paired events during Pavlovian conditioning and capable of reasoning causally about their interventions. These findings corroborate the evidence of Blaisdell et al. (2006) for these capacities for causal reasoning in rats, thereby strengthening the suggestion that this capacity, at least in rudimentary form, is shared with humans (see Meder et al., 2008; Waldmann & Hagmayer, 2005). The demonstrated competencies clearly pose a serious problem for the claim that causal reasoning in rats can be reduced to associative processes, and they weaken the argument that there is a sharp dividing line with respect to causal reasoning between human and nonhuman animals (e.g., Penn & Povinelli, 2007; Povinelli, 2000). Although our results suggest that capacities for many of the core features of causal cognition are present in rats, this of course does not imply that rats share all of the capacities that humans have. Causal cognition is not a unitary cognitive capacity but rather an amalgamation of many interacting processes and abilities. Obviously our study focuses on reasoning processes on the basis of observed covariations. The study does not demonstrate that causal reasoning is integrated into action planning, that rats can reason about physical mechanisms or complex relational causal structures (Penn, Holyoak, & Povinelli, 2008), or that they are capable of counterfactual reasoning (Sloman, 2005). These competencies need to be addressed in future research. Previous research, however, has often prematurely questioned the abilities of animals (and also of humans) by confronting them with overly complex tasks that do not clearly isolate the competence under study (e.g., causal reasoning) from other competencies that may also be needed to master the task (e.g., complex problem solving; see Waldmann & Walker, 2005). Complex problem solving (Funke & Frensch, 2007; Silva, Page, & Silva, 2005) and counterfactual reasoning (Meder et al., 2008) are tasks in which many humans fail without setting them apart from the rest of the human race.

Rather than looking for task differences, which are easy to find within and across different species, we should examine the underlying components of each task, many of which might be common across species. Through this approach we can move away from anthropocentric comparisons between tasks that human and nonhuman animals may or may not master and shift to the more fruitful focus on similarities and dissimilarities of cognitive components underlying different tasks.

References

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

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20,* 191–208.

Blaisdell, A. P. (2008). Cognitive dimension of operant learning. In J. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference: Vol. 1. Cognitive psychology of memory* (pp. 173–195). Oxford, England: Elsevier.

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

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006, February 17). Causal reasoning in rats. *Science, 311*(5763), 1020–1022.

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

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

Escobar, M., Arcediano, F., & Miller, R. R. (2001). Conditions favoring retroactive interference between antecedent events (cue competition) and between subsequent events (outcome competition). *Psychonomic Bulletin & Review, 8,* 691–697.

Escobar, M., & Miller, R. R. (2003). Timing in retroactive interference. *Learning & Behavior, 31,* 257–272.

Funke, J., & Frensch, P. A. (2007). Complex problem solving: The European perspective. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25–47). Hillsdale, NJ: Erlbaum.

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,* 3–32.

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

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Science, 9,* 290–295.

Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition, 12,* 695–707.

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

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 86–100). Oxford, England: Oxford University Press.

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

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

Killeen, P. R., & Smith, J. P. (1984). Perception of contingency in conditioning: Scalar timing, response bias, and erasure of memory by reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 333–345.

Lagnado, D. A., & Sloman, S. A. (2004). Inside and outside probability judgment. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 157–176). Malden, MA: Blackwell.

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

Limongelli, L., Boysen, S. T., & Visalberghi, E. (1995). Comprehension of cause-effect relations in a tool-using task by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 109,* 18–26.

Mackintosh, N. J. (1974). *The psychology of animal learning.* London, New York: Academic Press.

Matute, H., & Pineño, O. (1998). Stimulus competition in the absence of compound conditioning. *Animal Learning & Behavior, 26,* 3–14.

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

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

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

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–178.

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

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

Seed, A. M., Tebbich, S., Emery, N. J., & Clayton, N. S. (2006). Investigating physical cognition in rooks, *Corvus frugilegus. Current Biology, 16,* 697–701.

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.

Silva, F. J., Page, D. M., & Silva, K. M. (2005). Methodological-conceptual problems in the study of chimpanzees' folk physics: How studies with adult humans can help. *Learning & Behavior, 33,* 47–58.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., & Posada, A. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience, 7,* 80–84.

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

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

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

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

Underwood, B. (1966). *Experimental psychology*. Englewood Cliffs, NJ: Prentice Hall.

Visalberghi, E., & Limongelli, L. (1994). Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 108,* 15–22.

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, England: Oxford University Press.

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., & 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., & Walker, J. M. (2005). Competence and performance in causal learning. *Learning & Behavior, 33,* 211–229.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Winterbauer, N. E. (2006). *Associative and motivational bases of conditioned reinforcement.* Unpublished doctoral dissertation, University of California, Los Angeles.

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

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.

Received September 6, 2007
Revision received March 12, 2008
Accepted March 13, 2008 ■