Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review

Derek C. Penn and Daniel J. Povinelli

Cognitive Evolution Group, University of Louisiana, Lafayette, Louisiana 70504;
email: ceg@louisiana.edu

Key Words

causal reasoning, animal cognition, causal learning, contingency learning, causal Bayes nets, causal power, associationism

Abstract

In this article, we review some of the most provocative experimental results to have emerged from comparative labs in the past few years, starting with research focusing on contingency learning and finishing with experiments exploring nonhuman animals' understanding of causal-logical relations. Although the theoretical explanation for these results is often inchoate, a clear pattern nevertheless emerges. The comparative evidence does not fit comfortably into either the traditional associationist or inferential alternatives that have dominated comparative debate for many decades now. Indeed, the similarities and differences between human and nonhuman causal cognition seem to be much more multifarious than these dichotomous alternatives allow.

Contents

INTRODUCTION

Animals of all taxa have evolved cognitive mechanisms for taking advantage of causal regularities in the physical world. Many are also quite adept at using and manufacturing simple tools. But the way that human subjects cognize causal regularities is clearly a good deal more sophisticated than that of any other animal on the planet. This much, at least, seems indisputable. Which specific cognitive mechanisms human beings share with other animals, however, and which—if any— are uniquely human is an age-old question that is still very much unresolved (see, for example, Castro & Wasserman 2005, Chappell 2006, Clayton & Dickinson 2006, Reboul 2005, Vonk & Povinelli 2006).

In our opinion, substantive progress on this fundamental question has been greatly hindered by the dichotomous debate between "associationist" and "inferential" theories of causal cognition that has dominated comparative research for many decades and still holds sway in some parts of town (see Shanks 2006 for a review). Associationists have long argued that all of nonhuman causal cognition and most of human causal cognition as well can be reduced to a kind of contingency learning based on stimulus-bound, associative mechanisms similar to those that govern Pavlovian conditioning (for reviews, see Dickinson 2001, Pearce & Bouton 2001, Shanks 1995, Wasserman & Miller 1997). More generous comparative researchers, on the other hand, claim that even nonhuman animals are capable of reasoning about "causal-logical" relations in a human-like fashion (Call 2004) and employ "controlled and effortful inferential reasoning processes" to do so (Beckers et al. 2006).

In this article, we review some of the most provocative experimental results to have emerged from comparative labs in recent years. We start with experiments focusing on contingency learning and finish with research that explores nonhuman animals' understanding of unobservable causal mechanisms and causal-logical relations. The theoretical explanation for these provocative results is often inchoate at best. Nevertheless, a clear pattern emerges. The available comparative evidence does not fit comfortably into either the traditional associationist or classically inferential alternatives. Indeed, the similarities and differences between human and nonhuman causal cognition seem to be much more multifarious and fascinating than these

dichotomous alternatives allow (for similar suggestions, see Chappell 2006, Clayton & Dickinson 2006, Gallistel 2003, Heyes & Papineau 2005, Povinelli 2000, Shettleworth 1998).

LEARNING ASSOCIATIONS
Retrospective Revaluation Effects

We begin our bottom-up review of the recent comparative literature on causal cognition with a phenomenon that has challenged associationist theory from within. "Retrospective revaluation" is an umbrella term that associationist researchers use to refer to contingency learning scenarios in which the associative strength between a conditioned stimulus (CS) and an unconditioned stimulus (US) changes even though the CS in question is absent on the relevant training episodes. The paradigmatic example of a retrospective revaluation effect is backward blocking.

In the case of forward blocking, a cue is first paired with a US (e.g., A+) and then the first cue is presented in compound with a target cue and the US (e.g., AX+). Kamin (1969) showed that rats presented with A+ and then AX+ trials subsequently exhibited a weaker response to X than did rats who were only exposed to AX+ trials—as if learning that the reward was contingent on A "blocked" the rats from subsequently learning that the reward might also be contingent on X. The venerable Rescorla-Wagner (1972) model of associative learning was developed, in large part, to account for cue competition effects such as forward blocking.

In the case of backward blocking, the compound cue is trained first (AX+) and then the competing cue is presented alone (A+). As in forward blocking, the response to X alone is "blocked" on subsequent trials. This time, however, the blocking effect has occurred even though the X stimulus was absent on the critical A+ trials. There is compelling evidence for a variety of retrospective revaluation effects in nonhuman as well as human subjects (e.g., Balleine et al. 2005, Blaisdell &

Miller 2001, Denniston et al. 2003, Dickinson & Burke 1996, Miller & Matute 1996, Shanks 1985, Shevill & Hall 2004, Wasserman & Berglan 1998, Wasserman & Castro 2005).

Theoretical Accounts of Retrospective Revaluation Effects

Traditional associationist theories, such as the Rescorla-Wagner model, cannot account for retrospective revaluation effects because they assume that only cues present on a given trial can undergo a change in their response-eliciting potential (see discussions in Dickinson 2001, Shanks 2006, Wasserman & Castro 2005). Thus, there have been numerous attempts to revise associationist models in order to account for retrospective revaluation effects (e.g., Chapman 1991, Dickinson & Burke 1996, Van Hamme & Wasserman 1994). These revised associationist models postulate that subjects form within-compound associations between CSs that have occurred together in addition to associations between a CS and a US.

One common characteristic of these revised models is that the associative strength of an absent CS can be updated only if it has previously been associated with a CS that is actually present on the given trial. Recently, however, researchers have shown that the conditioned response to a US is also sensitive to the relation between stimuli that have never actually co-occurred but are only indirectly linked to each other through a web of intermediary associations (e.g., De Houwer & Beckers 2002a,b; Denniston et al. 2001, 2003; Macho & Burkart 2002). For example, Denniston et al. (2003) presented rats with AX+ trials and then with XY+ trials. Rats who subsequently received A- extinction trials responded less strongly to the Y cue than did rats who received no such extinction trials even though the A and Y stimuli never occurred together. As Wasserman & Castro (2005) point out, none of the revised associative models is able to account for higher-order effects such as these since the relevant cues never actually occurred together.

Within-compound associations: associations formed between cues that occur in close physical and/or temporal proximity as distinct from associations that form between cues and outcomes

To inferentially minded researchers, higher-order retrospective revaluation looks like it requires "higher-order reasoning processes" and "conscious propositional knowledge" (De Houwer et al. 2005). According to an inferential account, after the A-extinction trials, the rats in Denniston et al.'s (2003) experiment learned that A was not the true cause. Given this "propositional knowledge," the rats deduced that X, not A, was the actual causal stimulus in the AX+ pairing and then, by further deductive inference, that Y must not have been the true causal stimulus in the XY+ pairing (De Houwer et al. 2005).

A higher-order reasoning account of retrospective revaluation certainly provides one possible explanation for the rats' behavior and has undeniable appeal from a folk psychological point of view. However, it is not the only possible explanation. Denniston et al. (2001), for example, propose an alternative hypothesis that does not require propositional representations or inferential reasoning (see also Blaisdell et al. 1998; Denniston et al. 2003; Stout & Miller, manuscript submitted).

According to this "extended comparator hypothesis," the response to a given CS results from a comparison between the representation of the US directly activated by the target CS and the representation of the US indirectly activated by other CSs with which the target CS has been directly or indirectly associated in the past. For example, in Denniston et al.'s (2003) experiment, the extended comparator hypothesis suggests that the rats' response to Y was modulated by the second-order association between Y and A as well as the first-order associations between Y and X and between X and A (see Denniston et al. 2001 for a detailed exposition).

The extended comparator hypothesis is based firmly on traditional associative principles like spatio-temporal contiguity and "semantically transparent associations" (Fodor 2003, cited by Shanks 2006). However, contrary to traditional associationist theories, the comparator hypothesis posits that cues do not compete for associative strength when they are learned; rather, they compete for control over the subject's behavior when they are evaluated. Indeed, colloquially speaking, the extended comparator hypothesis proposes that a subject's response to a target cue is diminished by the extent to which it is able to "think" of an alternative cause or predictor of the outcome in question (Stout & Miller, manuscript submitted). In short, in order to explain the effects of higher-order retrospective revaluation, the extended comparator hypothesis posits the kind of performance-focused, structured information-processing capabilities that associationists have traditionally eschewed.

The Debate Between Associationist and Inferential Accounts

Which hypothesis best explains higher-order retrospective revaluation effects in human and nonhuman animals is a matter of vigorous debate (Aitken & Dickinson 2005; Beckers et al. 2005; De Houwer et al. 2005; Denniston et al. 2003; Melchers et al. 2004; Wasserman & Castro 2005). Unfortunately, only the associative side of the dispute has provided a formal specification of its claims. Van Overwalle & Timmermans (2001), for example, have proposed a connectionist implementation of Dickinson & Burke's (1996) model of first-order retrospective revaluation. And Stout & Miller (manuscript submitted) have submitted a formal computational specification of the extended comparator hypothesis. Higher-order reasoning accounts of retrospective revaluation, on the other hand, have only been formulated in a "verbal manner rather than formalized mathematically" (De Houwer et al. 2005). As De Houwer et al. (2005) frankly admit, "The most troubling implication of this lack of precision is that it becomes difficult to refute higher-order reasoning accounts." Obviously, an important future challenge for advocates of inferential accounts is to provide a formal, computational specification of their claims.

Regardless of which theoretical account prevails, the existing evidence has already demonstrated that nonhuman animals are capable of feats of causal learning once denied them by traditional associationist theory. Even cognitively minded researchers may need to revise their assessment of nonhuman causal cognition upwards. Visalberghi & Tomasello (1998), for example, once argued that nonhuman primates are unable to understand the "web of possibilities" that connects causes and effects. "Associative learning," Visalberghi & Tomasello (1998) went on to explain, "does not involve a web of possible connections, but only a one-to-one connection between antecedent and consequent." Evidence of higher-order retrospective revaluation in rats demonstrates that at least some nonhuman animals are, in fact, sensitive to higher-order associations between absent cues. And the extended comparator hypothesis demonstrates that the principles of associative learning can, in fact, be revised to take this web of possible connections into account.

ESTIMATING CAUSAL POWER

Ceiling Effects

If an effect, E, always occurs at its maximal level in a given context, regardless of whether a particular cause, C, is present or not, it is impossible to draw any inferences about whether the given cause has the power to produce the effect or not (Cheng 1997, Cheng & Holyoak 1995). The rational response to this state of affairs is for the subject to remain agnostic as to the causal power of the candidate cause in question. A subject interested in evaluating whether or not C prevents E, however, could infer that C is indeed noncausal. Ceiling effects are not symmetrical for generative and preventive causes (Cheng et al. 2006). It has now been well established that both human and nonhuman subjects are sensitive to ceiling effects when learning about contingen-

cies and treat generative and preventive cases differently (see Cheng 1997 for a detailed discussion).

Researchers have recently shown that both human and nonhuman subjects are also sensitive to outcome maximality and additivity effects (Beckers et al. 2005, 2006; De Houwer et al. 2002; Lovibond et al. 2003; Vandorpe et al. 2005). For example, Lovibond et al. (2003) showed that human subjects exhibit significantly stronger blocking when the candidate causes are described as having an additive effect. And Beckers et al. (2006) have shown that forward blocking in rats is attenuated when the intensity of the US presented during test trials is the same as the maximum intensity of the US experienced during prior training trials.

The Power PC Model of "Causal Power"

The traditional model for computing the statistical contingency between two cues is the ΔP model (Jenkins & Ward 1965, Rescorla 1968),

$$\Delta P = p(E|C) - p(E \mid \sim C),$$

where $p(E \mid C)$ is the probability of observing the effect given the presence of the candidate cause, and $p(E \mid \sim C)$ is the probability of observing the effect in the absence of the candidate cause.

The ΔP model of statistical contingency does not account for ceiling effects or the asymmetry between generative and preventive causes. Cheng (1997) showed that the normative model for estimating the contingency between a cause and an effect must consider the base rate probability of the effect in the absence of the candidate cause. In the case of binary causes and effects (i.e., causes and effects that are either present or absent), and assuming that the candidate cause is independent of any alternative causes in the subject's "focal set," Cheng's model for generative

Causal power: the probability with which a candidate cause, when present, actually produces or prevents an effect in question as distinct from the frequency with which the cause and the effect happen to co-occur

Preventive cause: a cause that has the power to prevent (or reduce the probability of) the occurrence of a given effect

Outcome maximality: the highest previous level at which a given outcome has been experienced by the subject

Candidate cause: of all the possible causes for a given effect, the one that is currently being evaluated by the subject

Base rate: the rate at which a given effect occurs in the absence of the candidate cause

Focal set: the set of events that a subject selects as relevant for computing the causal power of a candidate cause in a given context

causes is given by

$$q = \frac{\Delta P}{1 - p(E \mid \sim C)},$$

where ΔP is the standard model of statistical contingency described above, $p(E \mid \sim C)$ is the probability of observing the effect in the absence of the candidate cause, and q is an estimate of the unobservable "causal power" of the candidate cause in question.

In principle, the Power PC theory provides a normative model of causal induction in the absence of prior domain-specific causal knowledge that can handle both ceiling effects and the asymmetry between generative and preventive causes. It also handles retrospective revaluation effects by positing that subjects calculate causal power over the appropriate "focal set" of cases as specified by the theory (Cheng & Holyoak 1995).

To be sure, whether or not human causal intuitions actually conform to the Power PC model's predictions for intermediate base-rate probabilities is a matter of some dispute (Allan 2003, Buehner et al. 2003, Griffiths & Tenenbaum 2005, Lober & Shanks 2000, Perales & Shanks 2003). Worse, at least from a comparative point of view, there still have been no experiments testing the Power PC model's central predictions about the interaction between ΔP and base-rate probabilities on nonhuman subjects. In the absence of such evidence, it is still too soon for advocates of the Power PC theory to rest on their laurels.

Do Ceiling Effects Require an Inferential Explanation?

Regardless of whether or not the Power PC theory winds up being an appropriate model of causal induction for human or nonhuman subjects, the evidence for outcome maximality and additivity effects suggests that both human and nonhuman animals are not simply learning about observable contingencies but are sensitive to the unobservable constraints specific to causal inference. Both human and

nonhuman animals appear to understand tacitly that covariation only implies causation under special circumstances.

But does ruling out associationist explanations and acknowledging a profound similarity between human and nonhuman causal induction mean that rats necessarily employ "controlled and effortful inferential reasoning processes" and "conscious propositional knowledge" (Beckers et al. 2006, De Houwer et al. 2005)?

Beckers et al. (2006) do not present any computational arguments to justify their propositional attributions. Instead, they argue by analogy to human psychology: i.e., in human causal learning, sensitivity to outcome additivity and maximality seems to involve conscious propositional inferences, so nonhuman animals who exhibit a similar sensitivity must be employing similar mental processes (for an extended critique of this venerable argument, see Povinelli et al. 2000).

There are numerous reasons to be skeptical of this particular analogy. Almost all of the evidence cited in support of the role of conscious inferential processes in human causal cognition has no parallel in nonhuman studies. For example, Vandorpe et al. (2005) showed that the "verbal self-reports" of human subjects are consistent with an inferential account of blocking; blocking in human subjects is sensitive to secondary task difficulty (De Houwer & Beckers 2003, Vandorpe et al. 2005); and verbal information provided to human subjects after all learning trials have concluded nevertheless influences blocking effects (De Houwer 2002). These results certainly suggest a role for "controlled and effortful inferential reasoning" and "propositional knowledge" in human causal cognition. But there is no comparable evidence for nonhuman subjects. Moreover, such verbal reports in human beings may sometimes be just posthoc redescriptions of effects initially generated through implicit nonpropositional mechanisms (Povinelli et al. 2000). It is worth noting, in this respect, that Cheng (1997) has

consistently argued that subjects "implicitly" use a "qualitative" version of her model.

Beckers et al. (2005) do not present a formal specification of their hypothesis, so it is difficult to know exactly what they mean when they claim that outcome maximality and additivity effects in rats reflect the operation of "symbolic causal reasoning processes." Certainly, Beckers et al.'s (2005) evidence suggests that rats are able to manipulate representations that stand in for objective properties of stimuli in the world (e.g., the maximum intensity level of a given US) and are able to compute numeric operations over these values (e.g., updating the strength of a "blocked" cue as a function of the actual versus expected outcome). There are good reasons for believing that such information-processing operations require the ability to manipulate symbols and variables (see Marcus 2001 for a lucid discussion). Indeed, Buehner et al. (2003) make a similar claim with respect to the Power PC model. So if this is what Beckers et al. (2005) mean by "symbolic causal reasoning" processes, they seem to be on solid ground.

On the other hand, if what Beckers et al. (2005) mean by "symbolic causal reasoning" is equivalent to the kind of higher-order propositional inferences they attribute to humans, their claim is much more tendentious. In other domains—such as foraging, spatial cognition, and instrumental learning—comparative researchers have proposed a variety of information-processing architectures that could account for the observed outcome maximality and additivity effects in rats without positing the need for higher-order inferential reasoning or propositional representations (e.g., Clayton et al. 2001, Dickinson & Balleine 2000, Gallistel 2003, Shettleworth 1998). Given the species-specific evidence for conscious higher-order inferential reasoning in human cue competition effects (e.g., De Houwer et al. 2005), the suggestion that "parallel processes" are operating in humans and rodents (Beckers et al. 2006) appears premature.

INTERVENING ON CAUSAL STRUCTURES

Causal Bayes Nets

An increasingly influential approach to causal induction proposes that subjects solve causal reasoning problems in a manner consistent with the "causal Bayes net" formalism originally developed in computer science and statistics (Pearl 2000, Spirtes et al. 2001). The mathematical details of the causal Bayes net approach are far beyond the scope of the present review (but see Glymour 2003 for a short, nontechnical introduction). The basic idea, however, can be expressed easily enough in nonmathematical terms.

Causal Bayes nets represent causal structures as directed acyclic graphs in which nodes represent events or states of the world, and the connections between nodes (called "edges") represent causal relations. Some simple examples of possible causal structures include common-cause, common-effect, and causal chain models (see **Figure 1**).

The causal Bayes net formalism is predicated on a set of core assumptions. The most important of these assumptions is the causal Markov condition. The causal Markov condition says that if one holds all the direct causes of a given variable constant, then that variable will be statistically independent of all other variables in the causal graph that are not its effects. For example, in a simple common-cause model in which two effects occur with a certain probability given a particular value of C, the state of the two effects is independent of each other if the value of C is fixed.

The causal Bayes net formalism places special importance on the distinction between interventional and observational predictions (Danks 2006, Hagmayer et al. 2007, Pearl 2000, Spirtes et al. 2001, Woodward 2003). Interventions are formally modeled as an external independent cause that fixes the value of a given node, thereby "cutting off" all other causal influences on that manipulated node. Pearl (2000) has aptly baptized

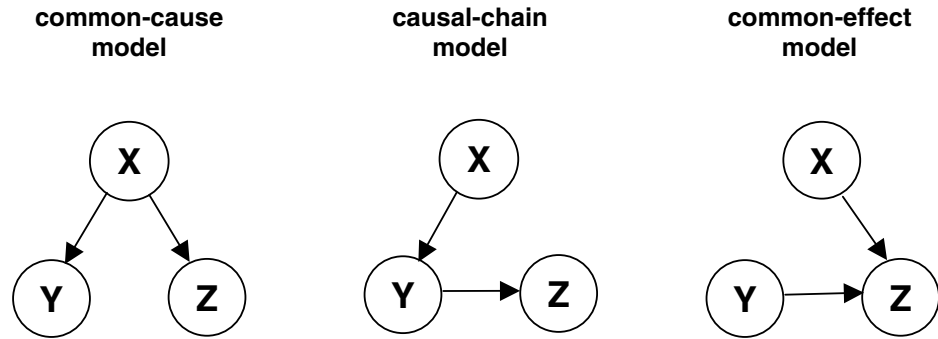## common-cause model   ## causal-chain model   ## common-effect model

Figure 1

Three basic causal structures represented as directed acyclic graphs. In a common-cause structure, X is the common cause of Y and Z. In a causal-chain structure, X influences Y, which influences Z. And in a common-effect structure, X and Y both influence a common effect, Z.

this procedure "graph surgery." Causal Bayes nets provide a formal, computational specification for how to derive interventional predictions from observational learning and vice-versa.

Different versions of the causal Bayes net approach propose different learning algorithms for inferring the causal structure underlying a given set of covariation data. Bottom-up models focus on providing algorithms for inferring causal structures based on statistical data in the absence of any other cues (e.g., Gopnik et al. 2004, Spirtes et al. 2001). The "causal model" approach emphasizes the role of top-down domain-general assumptions that constrain and inform the induction process (Waldmann 1996, Waldmann & Hagmayer 2001, Waldmann & Holyoak 1992). The "theory-based" approach focuses on the influence of domain-specific prior knowledge (Tenenbaum & Griffiths 2003, Tenenbaum et al. 2006).

The causal Bayes net formalism is arguably the most powerful formal account of human causal inference available today (see, for example, Danks 2005, Gopnik et al. 2004, Hagmayer et al. 2007, Lagnado et al. 2005, Tenenbaum & Griffiths 2003). Whether or not it is a psychologically accurate description of causal cognition in any nonhuman subject is another matter.

Intervention Versus Observation in a Nonhuman Subject

To our knowledge, only a single published experimental paper has explicitly claimed that nonhuman animals reason about causal relations in a manner consistent with a causal Bayes net approach. In the crucial experiment in this paper, Blaisdell et al. (2006) presented rats with stimuli whose conditional dependencies purportedly corresponded to one of two alternative causal structures. Rats presented with a common-cause model were given pairings of a light, L, followed by a tone, T, and, then separately, the same light, L, followed by a food reward, F. Rats presented with a causal-chain model were given pairings of T followed by L and then, separately, L followed by F.

During the test phase, each of the two groups of rats was divided randomly into one of two test conditions, and a lever that had not been previously present was inserted into the test chamber. Rats in condition intervene-T received a presentation of T each time they pressed the lever. Rats in condition observe-T observed presentations of T independently of their own actions on the lever. The experimenters recorded the number of nose pokes that the rats made into the magazine where F had been delivered during the training phase (there was no actual food in the magazine during the test phase).

The authors found that rats in condition intervene-T who had witnessed the common-cause model made fewer nose pokes than did rats in condition observe-T. In contrast, there was no significant difference between the intervene-T and observe-T conditions for rats in the causal-chain group. Based on these results, the authors concluded:

> Rats made correct inferences for instrumental actions on the basis of purely observational learning, and they correctly differentiated between common-cause models, causal-chains and direct causal links. These results contradict the view that causal learning in rats is solely driven by associative learning mechanisms, but they are consistent with causal Bayes net theories. The core competency of reasoning with causal models seems to be already in place in animals, even when elaborate physical knowledge may not yet be available. (Blaisdell et al. 2006)

Analysis of Blaisdell et al.'s (2006) Results

We are not surprised that nonhuman causal induction is consistent with at least some predictions of a causal Bayes net approach. How could it not be? The causal Bayes net formalism provides an exceptionally powerful *lingua franca* that can give posthoc explanations for nearly any nonpathological causal inference (Danks 2006). The question of whether human or nonhuman learning is consistent with a causal Bayes net approach is meaningless unless one specifies the particular model at stake. Thus, what is most intriguing about Blaisdell et al.'s results is the particular assumptions that one must make in order to claim that the rats' behavior is consistent with a causal Bayes net approach.

For example, during the initial training phase, the rats in the common-cause group were never presented with L followed by both T and F as would happen if both cues were actually effects of a common cause. Instead, the L → T pairings were perfectly negatively correlated with the L → F pairings: i.e., every instance of L was followed by either L or T but never both. In other words, rats purportedly presented with a common-cause model were, in fact, never shown covariation information consistent with a common-cause model. The fact that the rats, nevertheless, acted as if they had inferred a common-cause structure should give causal Bayes net enthusiasts reason to pause.

Adopting a simple common-cause model based solely on the observed data would violate the causal Markov condition since T and F are not independent conditional on the state of L. In order for the rats' behavior to be consistent with the causal Markov condition, one must posit that the rats had some sort of prior bias that influenced their causal judgments. One possibility is that the rats were working under the tacit assumption that simpler causal structures (e.g., a simple common-cause model) are more likely than complex causal structures (e.g., a common-cause model with inhibitory edges between effects). Such a prior bias would be consistent with those causal Bayes net approaches that allow for the influence of top-down, domain-general assumptions such as the causal model approach advocated by Waldmann and colleagues (Waldmann 1996, Waldmann & Hagmayer 2001, Waldmann & Holyoak 1992), but it would not be consistent with any causal Bayes net account that generates causal inferences in a purely bottom-up fashion based solely on observed covariations. To be sure, while Waldmann and colleagues have provided extensive evidence in support of their causal model hypothesis with respect to human subjects, the evidence for extending this hypothesis to rats is much more tenuous.

Blaisdell et al.'s (2006) own explanation for why the rats inferred a common-cause structure given the anomalous data is not based on Waldmann et al.'s causal model hypothesis. Instead, the authors proposed that rats who have observed L → T pairings and then observe an L → F pairing "conservatively treat the absent but expected events [i.e., T] as

possibly present but missed" (Blaisdell et al. 2006). Given the fact that the T stimulus was a highly salient tone or noise of 10 seconds in duration, the claim that the rats assumed they had somehow "missed" this cue cries out for further experimental corroboration.

Indeed, Blaisdell et al. (2006) provide very little evidence that rats inferred a common-cause model from the data using a causal Bayes net approach at all. Since the pairings were deterministic, the only information a subject could use to distinguish a common-cause structure from a causal-chain structure was the temporal ordering of the cues: i.e., T and F both appeared 10 seconds after L. While some causal Bayes net theorists have emphasized the importance of temporal information for causal learning (see Lagnado et al. 2005 for an example), the use of temporal information to distinguish between alternative causal structures is certainly not unique to the causal Bayes net formalism.

This said, the most critical finding in Blaisdell et al.'s (2006) experiment clearly undermines a traditional associationist account of the rats' behavior (see also Clayton & Dickinson 2006). Rats in the intervene-T condition of the common-cause group were less interested in F than rats in the observe-T condition. Blaisdell et al.'s provocative results challenge any theory of causal cognition that cannot explain how subjects derive novel interventional predictions from purely observational learning. But ruling out a traditional associationist explanation of the rats' behavior does not necessarily mean that rats tacitly cognize their own interventions in a human-like fashion or use the causal Markov condition to do so.

There is an abundance of evidence demonstrating that human subjects are able to use their own interventions in a deliberately epistemic fashion (Danks 2006; Hagmayer et al. 2007; Lagnado & Sloman 2002, 2004; Povinelli & Dunphy-Lelii 2001; Steyvers et al. 2003; Waldmann & Hagmayer 2005; Woodward 2003). Steyvers et al. (2003), for example, showed that human subjects do

not intervene randomly when the number of interventions they are allowed to make is constrained; instead, they choose their interventions in order to provide the most diagnostic test of their initial hypotheses. Indeed, human subjects seem to plan their interventions like quasi-experiments in order to eliminate confounds and distinguish between possible causal structures when observational data alone are ambiguous (Lagnado et al. 2005).

Blaisdell et al.'s (2006) results are consistent with Pearl's concept of graph surgery if one interprets this term as meaning nothing more than implicitly treating two associated events as independent once the subject has intervened on the consequent event (Michael Waldmann, personal communication). Crucially, however, Blaisdell et al.'s (2006) results do not provide any evidence that rats tacitly cognize their own interventions in an epistemic fashion, are sensitive to the causal Markov condition, or plan their own interventions in a quasi-experimental fashion to elucidate ambiguous causal relations as human subjects do. If what the rats did in the results reported by Blaisdell et al. (2006) counts as graph surgery, it is graph surgery by an accidental surgeon.

REASONING ABOUT CAUSAL MECHANISMS

The Role of "Intuitive Theories" in Human Causal Cognition

All of the theories reviewed to this point have focused on causal induction and the part played by domain-general causal assumptions. Researchers who focus on the problems of causal induction—whether from an associationist, causal power, or causal Bayes net perspective—have largely tabled any discussion of domain-specific prior knowledge. Many have explicitly stipulated that the nettlesome problem is outside the scope of their models (e.g., Cheng 1997, Dickinson 2001). Even those theorists who argue for the

importance of top-down knowledge have largely focused on tightly canalized, domain-general assumptions rather than learned domain-specific representations (e.g., Lagnado et al. 2005, Waldmann 1996).

Nevertheless, nearly everyone admits that prior domain-specific knowledge is an integral aspect of human causal cognition outside of the laboratory. Human subjects almost always use their prior domain-specific knowledge to evaluate novel causal relations rather than bootstrap their way up from observed covariation information and domain-general assumptions alone (Tenenbaum & Griffiths 2003, Tenenbaum et al. 2006). Indeed, human subjects often seek out and prefer information about underlying mechanisms rather than rely solely on information about covariation (Ahn et al. 1995).

Many cognitively minded researchers claim that human children possess abstract, coherent, rule-governed representations about the unobservable causal mechanisms at work in specific domains such as physics, biology, and psychology. Some theorists argue that this "core knowledge" is highly canalized due to heritable mechanisms (Carey 1985, Keil 1989, Spelke 1994), whereas others posit that it is largely learned on the fly (Gopnik & Meltzoff 1997). In either case, it is widely agreed that a child's causal knowledge is "theory-like" in the sense that it provides principled, allocentric, coherent, abstract explanations for the unobservable causal mechanisms that govern a given domain.

We use the term "intuitive theory" to refer to a subject's coherent domain-specific knowledge about unobservable causal mechanisms (Gopnik & Schulz 2004, Tenenbaum et al. 2006). By "unobservable," we mean that these causal mechanisms are based on the structural or functional relation between objects rather than on perceptually based exemplars (cf. Vonk & Povinelli 2006). One of the salient characteristics of unobservable causal mechanisms, such as gravity and support, is that they can be generalized freely to disparate concrete examples that share little to no perceptually based featural similarity

We have no doubt that intuitive theories about unobservable causal mechanisms play a formative role in human causal cognition. In the remainder of this article, we review the comparative evidence to ascertain to what extent this is true of nonhuman animals as well (see also Povinelli 2000, Vonk & Povinelli 2006).

Nonhuman Animals' Understanding of Tools, Support, and Gravity

To date, the strongest positive claims concerning nonhuman animals' intuitive theories about the physical world have come from a series of seminal experiments carried out by Hauser and colleagues on nonhuman primates' understanding of tools (Hauser 1997; Hauser et al. 1999, 2002a,b; Santos et al. 2003, 2006).

Hauser (1997) showed that adult tamarin monkeys reliably preferred cane-like tools whose shape is the same as a previously functional tool to cane-like tools with a novel shape but familiar color and texture. Based on these results, Hauser (1997) claimed that tamarin monkeys can distinguish causally relevant from causally irrelevant properties of a tool and thus possess a "functional concept of artifacts." Similar results have been documented for infant tamarin monkeys with minimal prior exposure to manipulable objects (Hauser et al. 2002a) as well as rhesus macaques, vervet monkeys, and lemurs (see Hauser & Santos 2006 for a review). Analogous results have been shown in the domain of food (Santos et al. 2001, 2002), where the sets of relevant and irrelevant features are reversed.

These results add to the growing body of evidence that nonhuman animals indeed do possess evolved domain-specific predispositions that bias how they perceive and manipulate objects in the world in the absence of observed covariation information or direct instrumental learning (see Shettleworth 1998

for a review). On the other hand, a heritable discriminative bias is not the same thing as an intuitive theory. Nothing in Hauser et al.'s results suggests that monkeys possess any insight into why one set of features is more relevant to tools than to food. Nor is there any evidence that their discriminatory biases are abstract, allocentric, or theory-like in the sense attributed to human children. All of Hauser's results to date are consistent with a more modest hypothesis; i.e., nonhuman primates are predisposed to perceive certain clusters of features as more salient than others when selecting among potential tools without understanding anything about the underlying causal mechanisms involved.

There is not simply an absence of evidence that nonhuman primates possess an intuitive theory about tools, there is also consistent evidence of an absence. Povinelli and colleagues have performed an extensive series of tests on chimpanzees' understanding of physical causal mechanisms (see Povinelli 2000). The general conclusion of these experiments is that when tool-use tasks are carefully construed to tease apart observable and unobservable relations, chimpanzees consistently focus solely on the observable relations and fail to cognize the unobservable causal mechanisms at stake (see also Vonk & Povinelli 2006).

In one of these experiments, for example, Povinelli (2000, Chapter 10) replicated Piaget's (1952) cloth-pulling experiment, in which subjects are asked to pull a piece of cloth toward them in order to obtain a reward that is lying on the cloth but out of reach. Hauser et al. (1999) had previously reported that tamarin monkeys successfully distinguished between rewards lying on or off the cloth and understood the support relation "at an abstract level, tolerating all featural transformations." Povinelli and colleagues, however, systematically varied the featural cues available to the chimpanzees and found that they were only sensitive to certain perceptual relations—such as the degree of surface contact between the cloth and the reward—and were insensitive to the actual structural rela-

tion casually relevant to obtaining the reward. In particular, the chimpanzees appeared to be oblivious to whether or not the cloth was actually supporting the reward as opposed to simply being in contact with it.

A series of seminal experiments by Visalberghi et al. provides further evidence for the absence of any intuitive theory among nonhuman primates about purely abstract causal mechanisms such as gravity and support (Limongelli et al. 1995, Visalberghi et al. 1995, Visalberghi & Limongelli 1994, Visalberghi & Trinca 1989). For example, Visalberghi & Limongelli (1994) tested capuchin monkeys' ability to retrieve a piece of food placed inside a transparent tube using a straight stick. In the middle of the tube, there was a highly visible hole with a small transparent cup attached. If the subject pushed the food over the hole, the food fell into the cup and was inaccessible. After about 90 trials, only one out of the four capuchin monkeys learned to push the food away from the hole; and even this one learned the correct behavior through trial and error. Worse, once Visalberghi et al. rotated the tube so that the trap-hole was now facing up and causally irrelevant, the only successful capuchin still persisted in treating the hole as if it needed to be avoided—making it obvious that even this subject did not understand the causal relation between the trap hole and the retrieval of the reward. By way of comparison, it should be noted that children as young as three years of age successfully solve the trap-tube task after only a few trials (Visalberghi & Tomasello 1998).

Povinelli (2000, Chapter 4) replicated Visalberghi's trap-tube setup with seven chimpanzees. Only a single chimp performed above chance on the normal trap tube condition. When tested on the inverted trap condition, this chimp—like the single successful capuchin in Visalberghi's original experiment—failed to take the position of the trap into account. More recently still, Santos et al. (2006) has replicated many of Povinelli's (2000) experiments with tamarin

and vervet monkeys and found convergent results. Based on these experiments, Santos et al. (2006) now conclude that nonhuman primates' comprehension of tools "is more limited than previously stated."

The failure to understand unobservable causal mechanisms such as support and gravity in an abstract fashion is not limited to nonhuman primates. Seed et al. (2006) recently presented eight rooks, a species of corvid, with a clever modification to the traditional trap-tube task. Seven out of eight rooks learned the initial version of the modified trap-tube task quite rapidly. Nevertheless, when presented with a series of transfer tasks in which the visual cues that predicted success in the initial task were absent or confounded, only one of the seven subjects passed. In a separate follow-up experiment (Tebbich et al. 2006), none of the rooks passed the transfer task.

Seed et al.'s (2006) results add to the growing evidence that corvids are quite adept at using stick-like tools, perhaps even more adept than nonhuman primates (see, for example, Chappell & Kacelnik 2002, 2004; Weir et al. 2002). But as Seed et al. (2006) point out, these results also suggest that rooks, like nonhuman primates, do not have a species-universal understanding of "unobservable causal properties" like gravity and support. Instead, they appear to solve tool-use problems based on the observable features of the task and evolved, task-specific expectations about what features are likely to be most salient (see also Chappell 2006 on the importance of interindividual differences).

The Case for Diagnostic Causal Reasoning in Nonhuman Apes

Until recently, there has been a consensus that nonhuman animals do not seek out diagnostic causal explanations (Povinelli 2000, Povinelli & Dunphy-Lelii 2001, Premack & Premack 1994, Visalberghi & Tomasello 1998). Breaking with this comparative consensus, Call (2004, 2005) has recently argued that nonhuman apes are, in fact, quite good at seeking out diagnostic causal explanations based on "causal-logical relations" and "quite bad at associating arbitrary stimuli and responses." These tendentious claims are based on a set of experiments testing apes' inferences about the location of food (Call 2004).

In these experiments, Call (2004) presented 4 bonobos, 12 chimpanzees, 6 orangutans, and 8 gorillas with two opaque cups, one of which was baited with a food reward. In the first experiment, the experimenter either showed the contents of both cups to the subjects or shook both of the cups. Unsurprisingly, subjects strongly preferred the cup in which they had seen the reward. Of the 27 original subjects, 9 also preferred the cup in which the shaking motion was associated with a noise.

In the second experiment, the 9 successful subjects from experiment 1 were retested in a condition in which only one of the two cups was shaken. In some trials, the baited cup was shaken; in other trials, the empty cup was shaken. Out of the 9 subjects, 3 were above chance on the crucial condition in which only the empty cup was shaken and the ape had to infer "by exclusion" that the food was in the other cup.

In subsequent experiments, Call (2004) tested a number of arbitrary noises (such as tapping on the cup or playing the recorded sound of a shaking noise) against the actual noise produced by shaking the cup. The apes chose the baited cup more frequently on "causal" conditions than on "arbitrary" ones. Based on this evidence, Call (2004) argued that the apes had understood the "causal-logical relation between the cup movement, the food, and the auditory cue" and understood that "the food causes the noise."

There are numerous problems with Call's interpretation. We do not know enough about the learning history of the subjects involved in these experiments to rule out the alternative hypothesis that they were simply responding on the basis of previously learned contingencies. It seems quite plausible, for example, that these captive apes had previously learned that

a shaking noise (N) combined with a shaking motion (M) is jointly indicative of a reward (NM+), whereas a shaking noise without a shaking motion (N-) or a shaking motion without a shaking noise (M-) is not. Positive patterning of this kind is a well-documented phenomenon in the animal conditioning literature (see Wasserman & Miller 1997 for a discussion). None of Call's numerous manipulations ruled out this obvious alternative hypothesis. As we discussed above, there is ample evidence that nonhuman animals often reason about causal relations in a fashion that is inexplicable in associationist terms. Ironically, though, a quite traditional associative explanation suffices in this particular case.

Indeed, Call's (2004) claim that the apes understood the contingencies in a "causal-logical" fashion appears to be refuted by Call's own results. In experiment 3, Call presented the apes with an empty "shaken silent cup" and an empty "rotated silent cup" (i.e., turned upside down and then right side up again). In experiments 1 and 2, the same shaking motion produced an audible rattling noise when the cup contained food. Nevertheless, the subjects strongly preferred the silent shaken cup to the silent rotated cup. If the apes had in fact understood the causal-logical relationship involved, they would have inferred that neither cup contained food and would have chosen randomly between the two cups or, if anything, would have preferred the rotated cup. Call provides no "causal-logical" explanation for why the apes would strongly prefer a shaken silent cup to a rotated silent cup. Once again, an explanation based on simple associative conditioning seems to fit the bill.

CONCLUSIONS, PROBLEMS, SUGGESTIONS

We hope that the comparative evidence we have reviewed over the course of this article has demonstrated why the venerable dichotomy between associationist and inferential explanations of nonhuman causal cognition is both specious and unproductive. We

agree with Chappell (2006): The real situation seems to be much more complicated, multifarious, and fascinating.

Many aspects of both human and nonhuman causal learning are parsimoniously explained in terms of some form of associative conditioning. On the other hand, both human and nonhuman animals are also sensitive to constraints specific to causal relations *sensu strictu*—such as ceiling effects and the asymmetry between generative and preventive causes. This implies that causal induction is not simply reducible to contingency learning in either human or nonhuman subjects. Even nonhuman animals employ cognitive mechanisms that distinguish between causality and covariation.

With respect to nonhuman animals' understanding of their own instrumental actions, the evidence again suggests that nonhuman causal cognition lies somewhere outside the associationist and inferential alternatives. On the one hand, nonhuman animals' capacity for flexible goal-directed actions suggests that they explicitly represent the causal relation between their own action and its consequences as well as the value of the expected outcome (see Dickinson & Balleine 2000). Moreover, Blaisdell et al.'s (2006) provocative results suggest that rats tacitly differentiate between the consequences of interventions on different kinds of causal structures. On the other hand, there is still no convincing evidence that nonhuman animals of any taxa seek out diagnostic explanations of anomalous causal relations or deliberately use their own interventions in order to elucidate ambiguous causal dependencies. For the moment, such diagnostic, inferential reasoning abilities appear to be uniquely human.

Most importantly, there appears to be a fundamental discontinuity between human and nonhuman animals when it comes to cognizing the unobservable causal mechanisms underlying a given task or state of affairs. While many species (including humans) have a tacit understanding that some events have the unobservable "power" to

cause other events (Cheng 1997), nonhuman animals' causal beliefs appear to be largely content-free; that is, their causal beliefs do not incorporate an abstract representation of the underlying generative mechanisms involved (Dickinson & Balleine 2000). Reasoning about the unobservable causal-logical relation between one particular causal belief and another appears to be a uniquely human trait.

This is not to say that all nonhuman animals are uniform in their cognitive abilities or that human subjects always reason in abstract, inferential terms. Evolution has clearly sculpted cognitive architectures to serve specific functions in specific species (Shettleworth 1998). The remarkable tool-using abilities of corvids stand as a stark reminder that the humanist notion of a *scala naturae* in causal cognition is wishful thinking. Our hypothesis is simply that in the panoply of animal cognition, the ability to reason about unobservable, domain-specific causal mechanisms in a causal-logical fashion is a specifically human specialization (see also Povinelli 2000, 2004; Vonk & Povinelli 2006).

Why is it that only human animals are able to acquire and use representations about unobservable causal mechanisms? Unfortunately, answering this comparative explanandum is hampered by the fact that our understanding of how intuitive theories work in human subjects is inchoate at best.

Causal Bayes net models are often touted as the way to reconcile bottom-up causal induction and top-down causal knowledge (Danks 2005, Gopnik et al. 2004, Gopnik & Schulz 2004). Upon closer inspection, however, it is clear that causal Bayes nets are not up to the task. As Tenenbaum et al. (2006) point out, any formal specification of intuitive causal theories of human causal cognition must be able to account for the hierarchical coherence among causal relations at various levels of abstraction. Lien & Cheng (2000), for example, showed that human subjects are more likely to judge a candidate cause to be genuinely causal if it is "hierarchically consistent" with their prior knowledge about superordinate causal

relations. Unfortunately, the edges and nodes of the causal Bayes net formalism are not "sufficiently expressive" to evaluate what makes a given concrete causal relation more or less coherent with superordinate mechanisms at different levels of abstraction (Tenenbaum et al. 2006).

In the absence of any well-established formal account of intuitive theories in human causal cognition, it is well nigh impossible to give a formal explanation for the discontinuity between human and nonhuman abilities in this area. Nevertheless, in the interests of provoking future research and debate, we close by proposing a preliminary hypothesis.

Our hypothesis is that abstract causal reasoning—i.e., causal cognition that involves reasoning about the relation between causal predicates at various levels of generality—is intimately bound up with the dynamics of analogical reasoning. It is well known that human subjects often learn about novel and unobservable causal relations by analogy to known and/or observable ones: The structure of the atom, for example, is often described by analogy to the solar system; electricity is conceived of as analogous to a flowing liquid; gravity is like a physical force. As Lien & Cheng (2000) suggest, the process of acquiring and predicating abstract causal relations seems akin to analogical inference; i.e., novel causal relations are often learned "by analogy" to known superordinate relations, and superordinate causal schemas are often learned by systematically abstracting out the functional elements common to superficially disparate causal regularities. If this hypothesis is right, computational models of analogical inference may provide the missing link between bottom-up and top-down processes in human causal cognition (French 2002, Gentner et al. 2001, Holyoak & Thagard 1997).

Our hypothesis provides a computational explanation for why human and nonhuman abilities differ so dramatically when it comes to reasoning about unobservable and abstract causal mechanisms. With the exception of a single unreplicated experiment on

one language-trained chimp (i.e., Gillan et al. 1981), there is no evidence that any nonhuman animal is capable of analogical reasoning. *Ex hypothesi*, the reason why only human subjects can reason about unobservable domain-specific causal mechanisms is because only humans have the representational architecture necessary to reason by analogy.

SUMMARY POINTS

1. The evidence suggests that nonhuman causal cognition is significantly more sophisticated than can be accounted for by traditional associationist theories. In particular, both human and nonhuman animals do not simply learn about observable contingencies; they appear to be sensitive to the unobservable constraints specific to causal inference.

2. On the other hand, there is a lack of compelling evidence that nonhuman animals pursue diagnostic explanations of anomalous causal relations or deliberately use their own interventions in order to elucidate ambiguous causal dependencies.

3. Nonhuman animals do not appear capable of the kinds of causal-logical inferences employed by human subjects when reasoning about abstract causal relations.

4. Nonhuman casual cognition is not well served by either the traditional associationist or classically inferential alternatives that have dominated comparative debate for many decades.

FUTURE ISSUES

1. Can the new generation of associationist models—such as the extended comparator hypothesis—account for the richness of nonhuman causal cognition without giving up the representational-level parsimony that has been the hallmark of traditional associationist theory?

2. Can advocates of a higher-order inferential account of causal cognition provide a formal, computational specification of their claims?

3. Are nonhuman animals sensitive to the effects of intermediate base-rate probabilities as predicted by the Power PC theory?

4. To what extent can nonhuman animals use their own interventions in a deliberately diagnostic manner? For example, can they use their interventions to elucidate ambiguous causal structures?

5. Can nonhuman animals infer the presence of hidden and/or unobservable causes? Are causal Bayes nets a useful formalism for describing the cognitive processes responsible for nonhuman causal cognition?

6. What representational-level limitations account for the inability of nonhuman animals to reason about unobservable causal relations?

7. What is the computational role of analogical inference in human causal cognition?

ACKNOWLEDGMENTS

LITERATURE CITED

Ahn W, Kalish CW, Medin DL, Gelman SA. 1995. The role of covariation versus mechanism information in causal attribution. *Cognition* 54:299–352

Aitken MR, Dickinson A. 2005. Simulations of a modified SOP model applied to retrospective revaluation of human causal learning. *Learn. Behav.* 33:147–59

Allan LG. 2003. Assessing Power PC. *Learn. Behav.* 32:192–204

Balleine BW, Espinet A, Gonzalez F. 2005. Perceptual learning enhances retrospective revaluation of conditioned flavor preferences in rats. *J. Exp. Psychol.: Anim. Behav. Process.* 31:341–50

Beckers T, De Houwer J, Pineno O, Miller RR. 2005. Outcome additivity and outcome maximality influence cue competition in human causal learning. *J. Exp. Psychol.: Learn. Mem. Cogn.* 31:238–49

Beckers T, Miller RR, De Houwer J, Urushihara K. 2006. Reasoning rats: forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *J. Exp. Psychol.: Gen.* 135:92–102

Blaisdell AP, Bristol AS, Gunther LM, Miller RR. 1998. Overshadowing and latent inhibition counteract each other: support for the comparator hypothesis. *J. Exp. Psychol.: Anim. Behav. Process.* 24:335–51

Blaisdell AP, Miller RR. 2001. Conditioned inhibition produced by extinction-mediated recovery from the relative stimulus validity effect: a test of acquisition and performance models of empirical retrospective revaluation. *J. Exp. Psychol.: Anim. Behav. Process.* 27:48–58

Blaisdell AP, Sawa K, Leising KJ, Waldmann MR. 2006. Causal reasoning in rats. *Science* 311:1020–22

Buehner MJ, Cheng PW, Clifford D. 2003. From covariation to causation: a test of the assumption of causal power. *J. Exp. Psychol.: Learn. Mem. Cogn.* 29:1119–40

Call J. 2004. Inferences about the location of food in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla*, and *Pongo pygmaeus*). *J. Comp. Psychol.* 118:232–41

Call J. 2006. Descartes' two errors: reason and reflection in the great apes. In *Rational Animals?*, ed. S Hurley, M Nudds, pp. 219–34. Oxford, UK: Oxford Univ. Press

Carey S. 1985. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press

Castro L, Wasserman EA. 2005. *Associative learning in animals and humans*. Presented at Online Conf. Causality, organized by Inst. Cogn. Sci., Lyon and Univ. Geneva. **http://www.interdisciplines.org/causality/**

Chapman GB. 1991. Trial order affects cue interaction in contingency judgment. *J. Exp. Psychol.: Learn. Mem. Cogn.* 17:837–54

Chappell J. 2006. Avian cognition: understanding tool use. *Curr. Biol.* 16:244–45

Chappell J, Kacelnik A. 2002. Tool selectivity in a nonprimate, the New Caledonian crow (*Corvus Moneduloides*). *Anim. Cogn.* 5:71–78

Chappell J, Kacelnik A. 2004. Selection of tool diameter by New Caledonian crows. *Anim. Cogn.* 7:121–27

This important paper shows that rats are sensitive to outcome additivity and maximality effects.

A provocative and seminal experiment suggesting that the causal Bayes net formalism may be appropriate for nonhuman as well as human subjects.

The author presents evidence that nonhuman apes are able to reason diagnostically in a "causal-logical" fashion.

Cheng PW. 1997. From covariation to causation: a causal power theory. *Psychol. Rev.* 104:367–405

Cheng PW, Holyoak KJ. 1995. Complex adaptive systems as intuitive statisticians: causality, contingency and prediction. In *Comparative Approaches to Cognition*, ed. JA Meyer, HL Roitblat, pp. 271–302. Cambridge, MA: MIT Press

Cheng PW, Novick LR, Liljeholm M, Ford C. 2006. Explaining four psychological asymmetries in causal reasoning: implications of causal assumptions for coherence. In *Topics in Contemporary Philosophy (Vol. 4): Explanation and Causation*, ed. M O'Rourke. Cambridge, MA: MIT Press

Clayton NS, Dickinson A. 2006. Rational rats. *Science* 9:472–74

Clayton NS, Griffiths DP, Emery NJ, Dickinson A. 2001. Elements of episodic-like memory in animals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356:1483–91

Danks D. 2006. Causal learning from observations and manipulations. In *Thinking with Data*, ed. M Lovett, P Shah. Mahwah, NJ: Erlbaum. In press

Danks D. 2005. The supposed competition between theories of human causal inference. *Philos. Psychol.* 2:259–72

De Houwer J. 2002. Forward blocking depends on retrospective inferences about the presence of the blocked cue during the elemental phase. *Mem. Cogn.* 30:24–33

De Houwer J, Beckers T. 2002a. Higher-order retrospective revaluation in human causal learning. *Q. J. Exp. Psychol. B* 55:137–51

De Houwer J, Beckers T. 2002b. Second-order backward blocking and unovershadowing in human causal learning. *Exp. Psychol.* 49:27–33

De Houwer J, Beckers T. 2003. Secondary task difficulty modulates forward blocking in human contingency learning. *Q. J. Exp. Psychol. B* 56:345–57

De Houwer J, Beckers T, Glautier S. 2002. Outcome and cue properties modulate blocking. *Q. J. Exp. Psychol. A* 55:965–85

De Houwer J, Beckers T, Vandorpe S. 2005. Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learn. Behav.* 33:239–49

Denniston JC, Savastano HI, Blaisdell AP, Miller RR. 2003. Cue competition as a retrieval deficit. *Learn. Motiv.* 34:1–31

Denniston JC, Savastano HI, Miller RR. 2001. The extended comparator hypothesis: learning by contiguity, responding by relative strength. In *Handbook of Contemporary Learning Theories*, ed. RR Mowrer, SB Klein, pp. 65–117. Mahwah, NJ: Erlbaum

Dickinson A. 2001. Casual learning: an associative analysis. *Q. J. Exp. Psychol.* 54B:3–25

Dickinson A, Balleine B. 2000. Causal cognition and goal-directed action. In *The Evolution of Cognition*, ed. CM Heyes, L Huber, pp. 185–204. Cambridge, MA: MIT Press

Dickinson A, Burke J. 1996. Within-compound associations mediate the retrospective revaluation of causality judgements. *Q. J. Exp. Psychol. B* 49:60–80

Fodor JA. 2003. *Hume Variations*. Oxford, UK: Oxford Univ. Press

French RM. 2002. The computational modeling of analogy-making. *Trends Cogn. Sci.* 6:200–5

Gallistel CR. 2003. Conditioning from an information processing perspective. *Behav. Process.* 61:1–13

Gentner D, Holyoak KJ, Kokinov BN, eds. 2001. *The Analogical Mind: Perspectives from Cognitive Sciences*. Cambridge, MA: MIT Press

Gillan DJ, Premack D, Woodruff G. 1981. Reasoning in the chimpanzee: I. Analogical reasoning. *J. Exp. Psychol.: Anim. Behav. Process.* 7:1–17

Glymour C. 2003. Learning, prediction and causal Bayes nets. *Trends Cogn. Sci.* 7:43–48

Gopnik A, Glymour C, Sobel D, Schulz L, Kushnir T, Danks D. 2004. A theory of causal learning in children: causal maps and Bayes nets. *Psychol. Rev.* 111:1–31

Gopnik A, Meltzoff AN. 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press

Gopnik A, Schulz L. 2004. Mechanisms of theory formation in young children. *Trends Cogn. Sci.* 8:371–77

Griffiths A, Tenenbaum JB. 2005. Structure and strength in causal induction. *Cogn. Psychol.* 51:334–84

Hagmayer Y, Sloman SA, Lagnado DA, Waldmann MR. 2007. Causal reasoning through intervention. In *Causal Learning: Psychology, Philosophy and Computation*, ed. A Gopnik, L Schulz. Oxford, UK: Oxford Univ. Press. In press

Hauser MD. 1997. Artifactual kinds and functional design features: what a primate understands without language. *Cognition* 64:285–308

Hauser MD, Kralik J, Botto-Mahan C. 1999. Problem solving and functional design features: experiments on cotton-top tamarins, *Saguinus oedipus oedipus*. *Anim. Behav.* 57:565–82

Hauser MD, Pearson H, Seelig D. 2002a. Ontogeny of tool use in cottontop tamarins, *Saguinus oedipus*: innate recognition of functionally relevant features. *Anim. Behav.* 64:299–311

Hauser MD, Santos LR. 2006. The evolutionary ancestry of our knowledge of tools: from percepts to concepts. In *Creations of the Mind: Essays on Artefacts and Their Representation*, ed. E Margolis, S Lawrence. Oxford, UK: Oxford Univ. Press. In press

Hauser MD, Santos LR, Spaepen GM, Pearson HE. 2002b. Problem solving, inhibition and domain-specific experience: experiments on cotton-top tamarins, *Saguinus oedipus*. *Anim. Behav.* 64:387–96

Heyes CM, Papineau D. 2005. Rational or associative? Imitation in Japanese quail. In *Rational Animals?*, ed. M Nudds, S Hurley, pp. 187–196. Oxford, UK: Oxford Univ. Press

Holyoak KJ, Thagard P. 1997. The analogical mind. *Am. Psychol.* 52:35–44

Jenkins HM, Ward WC. 1965. Judgment of contingency between responses and outcomes. *Psychol. Monog.* 79(Suppl. 1):1–17

Kamin LJ. 1969. Selection association and conditioning. In *Fundamental Issues in Associative Learning*, ed. NJ Mackintosh, pp. 42–64. Halifax, Nova Scotia: Dalhousie Univ. Press

Keil F. 1989. *Concepts, Kinds and Cognitive Development*. Cambridge, MA: MIT Press

Lagnado DA, Sloman SA. 2002. Learning causal structure. In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, ed. B Bel, I Marlien, pp. 560–65. Mahwah, NJ: Erlbaum

Lagnado DA, Sloman SA. 2004. The advantage of timely intervention. *J. Exp. Psychol.: Learn Mem. Cogn.* 30:856–76

Lagnado DA, Waldmann MR, Hagmayer Y, Sloman SA. 2007. Beyond covariation: cues to causal structure. In *Causal Learning: Psychology, Philosophy and Computation*, ed. A Gopnik, L Schultz. New York: Oxford Univ. Press

Lien Y, Cheng PW. 2000. Distinguishing genuine from spurious causes: a coherence hypothesis. *Cogn. Psychol.* 40:87–137

Limongelli L, Boysen ST, Visalberghi E. 1995. Comprehension of cause-effect relations in a tool-using task by chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 109:18–96

Lober K, Shanks DR. 2000. Is causal induction based on causal power? Critique of Cheng 1997. *Psychol. Rev.* 107:195–212

Lovibond PF, Been SL, Mitchell CJ, Bouton ME, Frohardt R. 2003. Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Mem. Cogn.* 31:133–42

Macho S, Burkart J. 2002. Recursive retrospective revaluation of causal judgments. *J. Exp. Psychol.: Learn. Mem. Cogn.* 28:1171–86

Marcus GF. 2001. *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge, MA: MIT Press

Melchers KG, Lachnit H, Shanks DR. 2004. Within compound associations in retrospective revaluation and in direct learning: a challenge for comparator theory. *Q. J. Exp. Psychol.* 57B:25–53

Miller RR, Matute H. 1996. Biological significance in forward and backward blocking: resolution of a discrepancy between animal conditioning and human causal judgment. *J. Exp. Psychol.: Anim. Behav. Process.* 18:251–64

Pearce JM, Bouton ME. 2001. Theories of associative learning in animals. *Annu. Rev. Psychol.* 52:111–39

Pearl J. 2000. *Causality*. London: Cambridge Univ. Press

Perales JC, Shanks DR. 2003. Normative and descriptive accounts of the influence of power and contingency on causal judgment. *Q. J. Exp. Psychol.* 56A:977–1007

Piaget J. 1952. *The Origins of Intelligence in Children*. New York: Norton

Povinelli DJ. 2000. *Folk Physics for Apes*. New York: Oxford Univ. Press

Povinelli DJ. 2004. Behind the apes' appearance: escaping anthropomorphism in the study of other minds. *Daedalus: J. Am. Acad. Arts Sci.* Winter, pp. 29–41

Povinelli DJ, Bering JM, Giambrone S. 2000. Toward a science of other minds: escaping the argument by analogy. *Cogn. Sci.* 24:509–41

Povinelli DJ, Dunphy-Lelii S. 2001. Do chimpanzees seek explanations? Preliminary comparative investigations. *Can. J. Exp. Psychol.* 55:185–93

Premack D, Premack AJ. 1994. Levels of causal understanding in chimpanzees and children. *Cognition* 50:347–62

Reboul A. 2005. *Similarities and differences between human and nonhuman causal cognition*. Presented at Online Conf. Causality, organized by Inst. Cogn. Sci., Lyon and Univ. Geneva. **http://www.interdisciplines.org/causality/**

Rescorla RA. 1968. Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.* 66:1–5

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, ed. AH Black, WF Prokasy, pp. 64–99. New York: Appleton-Century-Crofts

Santos LR, Hauser MD, Spelke ES. 2001. Recognition and categorization of biologically significant objects by rhesus monkeys (*Macaca mulatta*): the domain of food. *Cognition* 82:127–55

Santos LR, Hauser MD, Spelke ES. 2002. Domain-specific knowledge in human children and nonhuman primates: artifacts and food. In *The Cognitive Animal*, ed. M Bekoff, C Allen, GM Burghardt, pp. 269–81. Cambridge, MA: MIT Press

Santos LR, Miller CT, Hauser MD. 2003. Representing tools: how two nonhuman primate species distinguish between the functionally relevant and irrelevant features of a tool. *Anim. Cogn.* 6:269–81

Santos LR, Pearson H, Spaepen G, Tsao F, Hauser M. 2006. Probing the limits of tool competence: experiments with two nontool-using species (*Cercopithecus aethiops* and *Saguinus oedipus*). *Anim. Cogn.* 9:94–109

Seed AM, Tebbich S, Emery NJ, Clayton NS. 2006. Investigating physical cognition in rooks (*Corvus frugilegus*). *Curr. Biol.* 16:697–701

Shanks DR. 1985. Forward and backward blocking in human contingency judgment. *Q. J. Exp. Psychol.* 37B:1–21

Shanks DR. 1995. *The Psychology of Associative Learning*. London: Cambridge Univ. Press

Shanks DR. 2006. Associationism and cognition: human contingency learning at 25. *Q. J. Exp. Psychol.* 59:In press

Researchers present nontool-using monkeys with tasks designed to mirror those that had previously challenged chimpanzees.

The authors present an exceedingly clever twist on the traditional trap-tube task to seven rooks, a nontool-using species of corvid.

Shettleworth SJ. 1998. *Cognition, Evolution and Behavior*. New York: Oxford Univ. Press

Shevill I, Hall G. 2004. Retrospective revaluation effects in the conditioned suppression procedure. *Q. J. Exp. Psychol.* 57B:331–47

Spelke E. 1994. Initial knowledge: six suggestions. *Cognition* 50:433–47

Spirtes P, Glymour C, Scheines R, eds. 2001. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press

Steyvers M, Tenenbaum JB, Wagenmakers EJ, Blum B. 2003. Inferring causal networks from observations and interventions. *Cogn. Sci.* 27:453–89

Stout SC, Miller RR. 2006. Sometimes competing retrieval (SOCR): a formalization of the comparator hypothesis. *Psychol. Rev.* Manuscr. submitted

Tebbich S, Seed AM, Emery NJ, Clayton NS. 2006. Non-tool-using rooks (*Corvus frigilegus*) solve the trap-tube task. *Anim. Cogn.* In press

Tenenbaum JB, Griffiths TL. 2003. Theory-based causal induction. In *Advances in Neural Information Processing Systems*, ed. S Becker, S Thrun, K Obermayer, pp. 35–42. Cambridge, MA: MIT Press

Tenenbaum JB, Griffiths TL, Niyogi S. 2006. Intuitive theories as grammars for causal inference. In *Causal Learning: Psychology, Philosophy, and Computation*, ed. A Gopnik, L Schulz. Oxford, UK: Oxford Univ. Press. In press

Van Hamme LL, Wasserman EA. 1994. Cue competition in causality judgments: the role of nonpresentation of compound stimulus elements. *Learn. Motiv.* 25:127–51

Van Overwalle F, Timmermans B. 2001. Learning about an absent cause: discounting and augmentation of positively and independently related causes. In *Connectionist Models of Learning, Development and Evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop*, ed. RM French, JP Sougne, pp. 219–28. London: Springer-Verlag

Vandorpe S, De Houwer J, Beckers T. 2005. Further evidence for the role of inferential reasoning in forward blocking. *Mem. Cogn.* 33:1047–56

Visalberghi E, Fragaszy DM, Savage-Rumbaugh ES. 1995. Performance in a tool-using task by common chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), an orangutan (*Pongo pygmaeus*), and capuchin monkeys (*Cebus apella*). *J. Comp. Psychol.* 109:52–60

Visalberghi E, Limongelli L. 1994. Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). *J. Comp. Psychol.* 108:15–22

Visalberghi E, Tomasello M. 1998. Primate causal understanding in the physical and psychological domains. *Behav. Process.* 42:189–203

Visalberghi E, Trinca L. 1989. Tool use in capuchin monkeys: distinguishing between performance and understanding. *Primates* 30:511–21

Vonk J, Povinelli DJ. 2006. Similarity and difference in the conceptual systems of primates: the unobservability hypothesis. In *Comparative Cognition*, ed. T Zentall, EA Wasserman, pp. 363–87. Oxford, UK: Oxford Univ. Press

Waldmann MR. 1996. Knowledge-based causal induction. In *The Psychology of Learning and Motivation*, ed. DR Shanks, KJ Holyoak, DL Medin, pp. 47–88. San Diego, CA: Academic

Waldmann MR, Hagmayer Y. 2001. Estimating causal strength: the role of structural knowledge and processing effort. *Cognition* 82:27–58

Waldmann MR, Hagmayer Y. 2005. Seeing versus doing: two modes of accessing causal knowledge. *J. Exp. Psychol.: Learn. Mem. Cogn.* 31:216–27

Waldmann MR, Holyoak KJ. 1992. Predictive and diagnostic learning within causal models: asymmetries in cue competition. *J. Exp. Psychol.: Gen.* 121:222–36

Wasserman EA, Berglan LR. 1998. Backward blocking and recovery from overshadowing in human causal judgement: the role of within-compound associations. *Q. J. Exp. Psychol. B* 51:121–38

Wasserman EA, Castro L. 2005. Surprise and change: variations in the strength of present and absent cues in causal learning. *Learn. Behav.* 33:131–46

Wasserman EA, Miller RR. 1997. What's elementary about associative learning? *Annu. Rev. Psychol.* 48:573–607

Weir AAS, Chappell J, Kacelnik A. 2002. Shaping of hooks in New Caledonian crows. *Science* 297:981

Woodward J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford, UK: Oxford Univ. Press

Contents

Biological and Genetic Processes in Development

Development in Societal Context

Culture and Mental Health

Personality Disorders

Social Psychology of Attention, Control, and Automaticity

Inference, Person Perception, Attribution

Self and Identity

Social Development, Social Personality, Social Motivation, Social Emotion

Research Methodology

Indexes

Errata

An online log of corrections to *Annual Review of Psychology* chapters (if any, 1997 to the present) may be found at http://psych.annualreviews.org/errata.shtml