

## 21. The Psychology of Causal Perception and Reasoning

David Danks

### 1. Introduction

Causal beliefs and reasoning are deeply embedded in many parts of our cognition (Sloman 2005). We are clearly “causal cognizers,” as we easily and automatically (try to) learn the causal structure of the world, use causal knowledge to make decisions and predictions, generate explanations using our beliefs about the causal structure of the world, and use causal knowledge in many other ways. Because causal cognition is so ubiquitous, psychological research on it is itself an enormous topic, and literally hundreds of people have devoted entire careers to the study of it. As such, this chapter will necessarily be woefully incomplete. Each of the sections below (except perhaps section 4) could easily be expanded to an entire book, and this chapter must (by necessity) leave unaddressed some areas of psychological research that are plausibly relevant to causal cognition.<sup>i</sup>

Causal cognition can be divided into two rough categories: causal learning (sections 2-4) and causal reasoning (section 5). The former encompasses the processes by which we learn about causal relations in the world at both the type and token levels; the latter refers to the ways in which we use those causal beliefs to make further inferences, decisions, predictions, and so on. The two types of causal cognition are clearly connected to one another, but psychological research on each has proceeded relatively independently from the other. Causal learning itself can be divided into two distinct types: causal perception (section 2) and causal inference (section 3). Causal perception consists of the relatively automatic, relatively irresistible perception of certain sequences of events as involving causation. For example, if a nearby car alarm goes off

when I close my own car door, then I cannot help but perceive my own action as causing the alarm, even though I know that my action was not causally relevant. Causal inference, on the other hand, consists of higher-level causal learning that is based largely on statistical relationships. For example, I learn that one drug is better than another for pain relief by considering the relevant (statistical) history. There are historical and sociological reasons for this split in research on causal learning, but there are also apparent differences in phenomenology, behavior, and underlying neural bases. The precise connection between causal perception and inference is discussed in more detail in section 4. Finally, research on non-human animals (section 6) has in recent years helped us to understand better the nature of human causal cognition by revealing ways in which our causal cognition is similar to, and differs from, that of other animals.

## 2. Causal perception

Consider looking at a computer screen with a red square in the center, and a green square moving smoothly towards the center from the left side. Suppose further that the green square stops when it first “touches” (i.e., is contiguous with) the red square, and the red square begins moving to the right at the same speed. Nothing about this description, or (seemingly) the visual information, indicates anything about causation; this really is nothing more than a sequence of images on a computer screen. Nonetheless, when presented with a sequence of images such as these, almost everyone will immediately and spontaneously say that the green square *caused* the red square to move.<sup>ii</sup> No conscious thought or reasoning seems to be required, and very few observers can avoid believing that the one block caused the movement of the other. This canonical instance of causal perception--referred to as the *launching effect*--was first

systematically explored by Albert Michotte, and catalogued in his 1946 book *La perception de la causalité* (translated in 1963). Michotte conducted over one hundred studies investigating exactly when the launching effect does and does not arise spontaneously in observers. He showed, for example, that the standard launching effect does not arise if there is a spatial or temporal gap between the objects' movements: if the green square stops short of the red one, or the red one's movement occurs (noticeably) after the green square touches it, then one does not experience any perception of causality.

Michotte's findings were originally quite surprising, but methodological concerns about his results have essentially all been answered, and the basic phenomenon of causal perception is now widely accepted (White 1995). Even infants are widely thought to experience causal perceptions (Leslie 1982; 1984; Leslie and Keeble 1987; Oakes 1994; Oakes and Cohen 1990). Post-Michotte research on causal perception has largely aimed to determine: (i) the exact conditions under which causal perception occurs; and (ii) the constituent processes of causal perception. As an example of the first line of research, Scholl and colleagues have shown that causal perception depends on context, and not just the primary objects (Choi and Scholl 2004; Scholl and Nakayama 2002). Suppose that the green square (in the original example) moves over top of the red square and stops when it completely covers the red square, and the red square only starts moving once it is completely covered. In these cases, people do not normally experience any causal perception; rather, they usually experience one object passing smoothly over another and changing color spontaneously in the middle of the motion. If, however, an ordinary launching event occurs somewhere else on the screen at the same time, then the experience of this sequence changes. In this case, it is viewed causally as the green square launching the red. Whether an image sequence is perceived causally can thus change depending on seemingly

irrelevant events elsewhere in the visual field, and even infants are affected by these sorts of contextual changes (Newman *et al.* forthcoming).

Research on the component processes of causal perception has focused on its development, and its neuronal bases. Developmentally, six-month-old infants seem to perceive the basic launching effect stimuli in terms of causality, rather than “simpler” perceptual features such as contiguity or persistence (Leslie 1982; Leslie and Keeble 1987).<sup>iii</sup> Oakes and Cohen (1990) found that causal perception arises for more complex stimuli (e.g., unusual trajectories) only in ten-month-olds, but not six-month-olds (see also Oakes 1994). Infants younger than six-months-old attend to (i.e., perceive) only some of the spatial and temporal components of a launching event; they do not seem to have “full” causal perceptions (Cohen and Amsel 1998). Causal perception thus seems to require (separable) perceptions of appropriate spatial and temporal contiguity, and also the ability to ignore extraneous perceptual elements. Causal perception does not arise in one fell swoop, but rather comes together more slowly. In adults, spatial contiguity even ceases to be necessary in some cases; causal perceptions can arise without any spatial contact at all between the on-screen objects (White and Milne 1997; 1999). Similar results emerge from the limited neuroscientific work on causal perception: for example, fMRI data suggest that overlapping, but not identical, brain regions are responsible for the spatial and temporal components of causal perception (Fugelsang *et al.* 2005). There does not seem to be a single, neuronally distinct “module” for causal perception (though see below).

Most experiments on causal perception have focused on variants of launching events, but causal perception can arise in other contexts. In their classic experiments, Heider and Simmel (1944) showed that the movement of simple geometric objects is sometimes perceived as *intentional* movement by the objects. For example, suppose the red square (in the original

launching event) begins moving before the green square arrives, and the red square moves erratically while the green square follows smoothly behind it. This sequence of images will typically be perceived as the red square “fleeing” while the green square “chases.” That is, causal perceptions arise not just for physical causation, but also for social or intentional causation, and are similarly automatic and unprompted in the latter domain. These perceptions of objects as “intentional agents” whose states can cause behavior seem to arise as early as nine-months-old (Csibra *et al.* 1999; Gergely *et al.* 1995). Physical and social causal perceptions do appear to be separable, however, as the former seems to be perceived more strongly than the latter (Schlottmann *et al.* 2006).

Causal perception has traditionally been viewed as philosophically interesting because it seems to be a (partial) psychological vindication of Kant over Hume: certain judgments of causality seem to be part-and-parcel of perception, rather than something that occurs after “basic” perception has taken place. Moreover, these causal perceptions can influence other perceptual judgments, such as event timing (Choi and Scholl 2006; Newman *et al.* forthcoming), and causal perception does not seem to be susceptible to top-down control or overriding (Blakemore *et al.* 2001; Fonlupt 2003). Causality seems to be built-in to some of our perceptions of the world, rather than always being only inferred from a sequence of images. More generally, as Michotte (1963) himself realized, causal perception seems to be a plausible candidate for a modular process (in the Fodor 1983 sense), as it is fast, automatic, mandatory, and informationally encapsulated. Causal perception (seemingly) depends only on visual input, and not on higher-level cognition; you cannot, for example, choose to not see the classic launching events as causal. It also arises cross-culturally, and causal perceptions are the same even for groups that make quite different causal attributions in social contexts (Morris and Peng 1994).

Several researchers have used the above reasons to argue that causal perception is plausibly a cognitive module (Leslie 1984; 1994; Leslie and Keeble 1987; Scholl and Tremoulet 2000), and perhaps even a neurological module (Blakemore *et al.* 2001). But although causal perception behaves modularly in processing, there are reasons to doubt that it constitutes a fully Fodorian module. It does not have a classically modular development (Schlottmann 2000), as causal perception requires different cognitive components that develop at different times. Neurally, these components seem to be distributed relatively widely: both temporal lobes, the inferior parietal lobe, and the frontal gyri (Blakemore *et al.* 2001; Fugelsang *et al.* 2005). Behaviorally, there do not seem to be any reported cases of selective loss of causal perception; at the current time, no individuals have been found with lesions or other neural damage that resulted in loss of causal perception (without much broader loss of visual perception). There are also significant individual differences in causal perception. Examples include the findings that (a) some individuals fail to have a causal perception for classic launching stimuli (e.g., Beasley 1968); (b) causal perceptions or their absence can change upon repeated exposure of the same stimuli; and (c) experience can affect whether causal perceptions occur. Moreover, these individual variations are largely stable over time (Schlottmann and Anderson 1993), and so suggest that the “module” at least has important parameters that are set by personal experience. Causal perception has many modular features--automaticity, mandatory triggering, and informational encapsulation--but does not seem to satisfy fully the classical profile of a module.

### 3. Causal inference

A different type of causal learning occurs when one is learning that exposure to a particular plant (e.g., poison ivy) causes a rash, or that a new drug has various side effects. In

these cases, one often cannot rely on spatiotemporal cues, but rather must attend to differences in occurrence rates in some relevant population. The paradigmatic situation for causal inference is one in which the learner observes a series of situations or cases in which various potential causes do or do not occur, and the presumptive effect does or does not occur (Cheng 1997; Cheng and Novick 1990; Shanks 1995). The learning challenge is then to determine (a) which potential causes are actual causes, and (b) the strengths (in some sense) of those causes. Numerous variations on this paradigmatic situation are obviously possible; for example, one might have spatiotemporal information (e.g., Buehner and May 2002), or the learner might actively bring about some of the cases (e.g., Sobel and Kushnir 2006). The central challenge remains largely the same, however: use principally statistical information (e.g., something like correlation) to learn causal relations and strengths. This type of causal inference is not directly perceptual, nor does it seem to have the same type of automaticity as causal perception: people rarely learn that a plant causes a rash after only one exposure (though they might suspect that it does so). Causal inference also seems to require higher-order (in some sense) cognition than causal perception. As a result, psychological research on causal inference has proceeded relatively independently of research on causal perception (though see section 4 below).

The dominant experimental paradigm in psychological research on causal inference has three principal components. First, the “cover stories” largely prevent experimental participants from using any substantive prior causal knowledge beyond, e.g., temporal order. Second, the relevant variables for causal inference are always obvious in the stimuli, and the variables might even be divided into potential causes and an effect. Third, participants provide their judgments about the causal relations as explicit numeric ratings for each potential cause, usually on a scale ranging from -100 (“always prevents”) to +100 (“always generates”). There are of course

experiments without one or another of these components: for example, the cover story might evoke substantive domain knowledge (e.g., Schulz and Gopnik 2004); the cases might be presented using actual objects (e.g., Gopnik *et al.* 2004); participant behavior might generate the cases (e.g., Buehner and May 2003; Steyvers *et al.* 2003); or participants might respond with graphs rather than numeric ratings (e.g., Steyvers *et al.* 2003). The principal theoretical task, however, is surprisingly constant over all of these variations: explain the patterns of ratings that are generated by systematic variations in the statistical relationship between the potential causes and effect.

One intuition explored early in the psychological research is that human causal inference might be similar to, or even identical to, the associative learning processes found in non-human animals. Most people are familiar with the notion of classical (Pavlovian) conditioning: repeatedly ring a bell (referred to as the cue or Conditioned Stimulus, CS) just before presenting a dog with food (the outcome or Unconditioned Stimulus, US) and the dog will come to associate bell-ringing with food (and so salivate upon bell-ringing). Instrumental conditioning refers to situations in which the relevant cue is generated through the animal's own action (e.g., the dog presses a lever). Of course, both types of conditioning can lead to quite complex patterns of behavior, as numerous behaviorist experiments demonstrated (e.g., Skinner's pigeons that famously "played" Ping Pong). Broadly speaking, formal models of these processes, and the full range of conditioning phenomena, are referred to as *associative* models (though in recent years, 'associative' has frequently been used as a pejorative term to refer to any model that an author does not like). The dominant associative model of the past thirty years is the Rescorla-Wagner (1972) model, though many alternatives have been proposed (e.g., Pearce 1987; Van Hamme and Wasserman 1994). All of the models represent associative learning as the learning of so-called

associative strengths for the different factors, and they share other features: the processes require little memory or computational power; cases are handled sequentially, rather than as a group; and learning proceeds through an error-correction process (i.e., associative strengths are adjusted based on the error between (a) their prediction about whether the outcome will occur, and (b) whether it actually does occur). There are also models that share these features, though they have no history in the animal behavior literature (Catena, Maldonado and Candido 1998; Danks, Griffiths and Tenenbaum 2003). One proposal for causal inference is that the causal strengths learned in human causal inference (and reported as ratings in experiments) might actually be associative strengths learned using some associative process (Shanks 1995).

Instead of the case-by-case learning modeled by associative models, one could focus on asymptotic causal inference: what causal relations do people learn after a long enough sequence or summary of cases (i.e., once their beliefs stabilize)? This type of causal inference is closely connected to, but importantly different from, the narrower problem of contingency learning: people's ability to infer association or independence between two variables given either a sequence of cases, or a summary table of the data (De Houwer and Beckers 2002 review empirical data on contingency learning; McKenzie and Mikkelsen 2007 review formal models). Three different models of asymptotic causal inference give a feel for their diversity. The  $\Delta P$  model (Cheng and Novick 1990; 1992) holds that causal strength judgments are given by the difference between the probability of the effect when (a) the potential cause is present, and (b) when it is absent (i.e.,  $\Delta P = P(E|C) - P(E|\neg C)$ ). The causal power approach (Cheng 1997; Novick and Cheng 2004) supposes that people represent the world in terms of unobserved causal powers (similar to capacities in the sense of Cartwright 1989) and use the observed statistics to try to make inferences about the strengths of those powers. The pCI model (White 2003a; c) argues

that people attend to the proportion of confirming instances: the fraction of observed cases that support the existence of a causal relation (relative to the total number of observed cases). There are differences in metaphysical commitments and mathematics, but the models of asymptotic causal inference all share the common goal of predicting people's rating patterns once learning is completed and beliefs have stabilized.

Associative models of causal inference and models of asymptotic causal inference are often thought to be direct competitors with one another. A series of formal results (e.g., Cheng 1997; Danks 2003; Tenenbaum and Griffiths 2001) have emerged in the past ten years, however, showing systematic connections between associative models of causal inference and models of asymptotic causal inference. Specifically, many different associative models of case-by-case learning each (provably) converge in the limit to a different asymptotic model (e.g., the associative model of Danks *et al.* 2003 converges to causal power). Danks (2007b) extends and unifies these disparate results, and shows that these different types of models do not compete, but rather are simply models at different temporal scales.

These mathematical connections reveal that both associative and asymptotic models focus on inference of causal strengths, rather than causal structure. Of course, strength ratings implicitly encode structure (i.e., no causal connection if and only if strength of zero), but the two types of inference are at least logically separable. The causal Bayes net framework (Pearl 2000; Spirtes, Glymour and Scheines 1993) explicitly represents this distinction between causal structure (i.e., the graph) and causal strength (i.e., the parameters). Moreover, all of the previously proposed causal inference theories--both associative and asymptotic--provably correspond to strength inference rather than structure inference (Danks 2007b; Griffiths and Tenenbaum 2005). The causal Bayes net framework also provides a clear account of the

difference between learning from observations and learning from interventions, which had previously been largely neglected in the psychological literature. That account led to numerous studies that confirmed that the observation vs. intervention difference affects people's causal inference (e.g., Gopnik *et al.* 2004; Lagnado and Sloman 2004; Sobel and Kushnir 2006; Steyvers *et al.* 2003).

This representational power, as well as the successful use of causal Bayes nets in other domains, has prompted two types of proposals for human causal inference based on causal Bayes nets. The first type holds that human causal inference involves learning causal structure and strengths from the set of all possible structures consistent with background knowledge (Gopnik and Glymour 2002; Gopnik *et al.* 2004; Griffiths and Tenenbaum 2005; Steyvers *et al.* 2003). These proposals differ principally about the nature and level of the learning algorithm. The second type argues that people start with some initial structure, and only change their mind if the data directly contradict the initial model (Hagmayer *et al.* 2007; Lagnado and Sloman 2004; Lagnado *et al.* 2007; Waldmann 1996). The initial structure is selected on the basis of various heuristics, such as “if I change  $X$ 's value, then anything that changes afterwards must be an effect of  $X$ .” Learning on this account does not involve selecting the best (by some measure) causal Bayes net from the set of all plausible possibilities; instead, the learner selects an hypothesis by various heuristics, and then retains it until it is falsified.

There are many different psychological models for causal inference, and a correspondingly large number of experimental studies. There is very little agreement about which causal inference model is right, or even about which approach is the most likely to be fruitful. One significant problem is that none of the extant models can capture all of the extant data (Perales and Shanks 2007). Some models are of course better than others, but none predict

all of the ways that ratings vary as statistical information changes. Causal Bayes net theories have a representational advantage in capturing the psychologically significant distinction between observation and intervention, but even they cannot explain all data relating to that distinction. One potential explanation is that people use a mixture of strategies (Buehner, Cheng and Clifford 2003; Lober and Shanks 2000), and different experiments may well elicit different types of judgments, as ratings seem to be sensitive to the probe question used (Collins and Shanks 2006; White 2003b). One might hope that neuroscience could help, but currently available fMRI data do not provide much insight. Causal inference does seem to involve some sort of error prediction/correction occurring principally in the prefrontal cortex (Corlett *et al.* 2004; Fletcher *et al.* 2001; Turner *et al.* 2004), but the data tell us nothing about how that calculation figures in causal inference. Semantic retrieval of causal information and semantic retrieval of associative information lead to different patterns of neural activation (Satpute *et al.* 2005), but again the data do not illuminate the nature of that difference.

The preceding discussion has largely focused on a limited subset of causal inference: principally, causal inference from statistical information, rather than other information. In practice, causal inference is sensitive to many other factors, such as knowledge of temporal features of the possible causal relations (Buehner and May 2002; 2003; Lagnado and Sloman 2004). More generally, causal inference is clearly influenced by prior beliefs. Strong covariations in observed data are more meaningful (i.e., lead to larger ratings) if people know a plausible mechanism underlying the covariation, rather than an implausible one; relatively little effect of mechanism plausibility in prior belief occurs for weak covariations, perhaps because ratings are already quite low (Fugelsang and Thompson 2003). This effect also seems to have a neural basis. Consistency between prior belief and observed data (i.e., plausible mechanism and strong

covariation, or implausible mechanism and weak covariation) activates learning and memory regions of the brain, while belief-data inconsistency activates error-correction and conflict resolution areas (Fugelsang and Dunbar 2005). Most generally, causal inference is significantly influenced by the categories and concepts that we have (Waldmann and Hagmayer 2006). People typically do causal inference with the categories that they have prior to learning, even when those categories are suboptimal for causal inference.

#### 4. Intersections between causal perception and causal inference

An obvious issue centers on the relationship, if any, between causal perception and causal inference. Are the processes identical? Is one necessary for the other? Is one a subset of the other? Are they entirely distinct? One way to (try to) address this issue is through developmental progressions; e.g., if one type of cognition appears before the other, then the later process presumably cannot be necessary for the earlier one. As noted earlier, causal perception has been found in six-month-olds. It is at least *prima facie* possible that infants of a similar age make causal inferences, particularly since at least eight-month-olds are sensitive to some statistical patterns in their environment (Saffran, Aslin and Newport 1996; Saffran *et al.* 1999). We do not know, however, the earliest age of causal inference, largely because there are obvious methodological challenges. There will typically be many alternative explanations for data from looking-time studies that suggest causal inference, such as the infants simply noticing predictively useful associations. We thus do not currently know whether one process emerges before the other.

A different approach is to explore whether the processes can be separated in adult cognition. Surprisingly, there has been relatively little research directly on this topic. On the

neurological front, Roser *et al.* (2005) examined causal perception and inference in two corpus callosotomy<sup>iv</sup> patients, and found that causal perception and causal inference seem to occur in different brain hemispheres (perception in right hemisphere, inference in left hemisphere). Independent fMRI studies on each type of causal learning also suggest a neuroanatomical difference. Causal perception seems to be concentrated in the temporal lobes (Blakemore *et al.* 2001; Fugelsang *et al.* 2005), while at least one significant part of causal inference--namely, error prediction and correction--seems to be largely localized in prefrontal cortex (Corlett *et al.* 2004; Fletcher *et al.* 2001; Turner *et al.* 2004). However, despite these apparent neuroanatomical differences between causal inference and perception, there are no known cases of selective loss of only one of the types of cognition. The neurological evidence is thus relatively ambiguous: causal perception and causal inference seem to occur at least partially in different brain regions, but it is unknown whether they are fully dissociable.

Yet another approach is to try to find situations that prompt only causal perception or only causal inference. Schlottmann and Shanks (1992) presented experimental participants with many different sets of launching-type sequences. In one set of sequences, spatiotemporal contact reliably led to movement of the “launched” block only after a noticeable delay. Participants came to recognize, presumably via causal inference, that the “launching” block was a cause, but they reported that “it just did not look as if it should be” (p. 338) and so gave relatively low causal perception ratings. In a second set of sequences, spatiotemporal contact was uncorrelated with subsequent launching, and color change in the launched block was instead the reliable predictor. The most interesting case for these sequences is when the second block moves after spatiotemporal contact, even though such contact is (overall) uncorrelated with movement. In this case, participants give high causal perception ratings for the “launching” block as a cause,

even though they recognize that it is entirely unnecessary; they report that the collision “just looked as if it should be” a cause (p. 338). This distinction between causal perception and causal inference is found in both participant ratings and phenomenological experiences (p. 339), which supports the idea that these types of cognition are actually separable.

Interactions between causal perception and causal inference can become quite complicated when a causal mechanism requires some time to operate. For example, suppose a button press causes a light to illuminate only (and always) after a three-second interval. Now suppose that one presses that button, and then presses it again three seconds later (when the light comes on). Causal perception says the second button press is the cause because of temporal contiguity; causal inference (or reasoning) says the first button press is the cause. Adults are largely able to use mechanism information to override the causal perception in these cases, but seven-year-old children are not (Schlottmann 1999). More generally, adult causal inference is influenced by knowledge of the timing of underlying mechanisms, as cause-effect relationships can be inferred even when there is a significant temporal gap between the two (Buehner and May 2002; 2003). Adult causal perception is not influenced by that knowledge, however, as spatiotemporally contiguous events are (almost) always perceived causally while separated ones are not. Explicit timing knowledge in causal inference can also shape which events are thought to be possible causes in the first place (Hagmayer and Waldmann 2002), but has no such impact on causal perception.

In summary, there is a growing body of direct and indirect evidence that causal perception and causal inference are different cognitive processes. The current psychological evidence does not, however, provide much information about the relationship between these processes. In particular, it is simply unknown whether one is necessary for the other--either

developmentally or cognitively--or they are (relatively) autonomous cognitive processes. One barrier to fruitful psychological research has arguably been the lack of understanding of the relevant theoretical “possibility space.” The space of possible relationships between causal perception and causal inference is largely unknown, and philosophical thought could potentially provide significant guidance in the development and testing of psychological theories.

## 5. Causal reasoning

Causal reasoning is also a significant part of causal cognition, and perhaps even constitutes the majority of adult causal cognition. One great value of causal knowledge is the myriad ways that we can use it to understand, predict, and control the world around us (Sloman 2005). Psychological research on human causal reasoning has historically been conducted in particular domains of application of the causal knowledge (e.g., decision making, categorization). In recent years, the causal Bayes net formalism has provided a small measure of unification to the reasoning research, but the work still largely consists of various disjoint research endeavors.

One commonplace type of causal reasoning is the use of causal knowledge to make decisions. Suppose I know (or believe) that  $X$  causes  $Y$ . If I desire  $Y$ , then I might naturally decide to try to bring about  $X$ . In contrast, if I desire  $X$ , then there is no particular value to bringing about  $Y$  directly. People are sensitive to this distinction, and they exhibit appropriate behavior whether they are taught the causal structure explicitly, or learn it from observed data (Hagmayer and Sloman 2005; Nichols and Danks 2007). People also appear to systematically treat their own decisions as occurring outside of the causal system; they act (except in very unusual situations) as though their decisions are uncaused by variables in the causal structure.

Moreover, many of the experiments used to explore causal inference employ behavioral measures of learning that depend on people's ability to do causal reasoning. For example, children are shown that some combinations of blocks ("blickets") activate a machine, and then their causal knowledge is assessed by asking them to make the machine go or stop (Gopnik *et al.* 2004). This behavioral measure--which block the child places on the detector, or removes from it--is discriminative of causal learning only if the children are able to use the products of learning to make decisions. This range of findings about decision-making based on causal reasoning have led to a formal model of decision-making (given causal beliefs) that is based on causal Bayes nets (Sloman and Hagmayer 2006), and experimental tests are ongoing.

Causal reasoning also occurs in the context of conceptual reasoning. One example of the relevance of causal beliefs to categories comes from the so-called causal status effect. If category *A* is partially characterized by the (possibly indeterministic) causal relation  $X \rightarrow Y$ , then individuals with *X* but not *Y* are systematically judged as more likely to be in *A* than individuals with *Y* but not *X* (Ahn *et al.* 2000; Rehder and Kim 2006). That is, if all *A*'s have *X* causing *Y*, then *X* is more important than *Y* in deciding whether some new individual is an *A*. Rehder and colleagues have argued that the connection between causal reasoning and concepts might be substantially deeper. In the past forty years, psychological theories of concepts have usually understood concepts in terms of observed features, whether prototypical instances (Posner and Keele 1968), sets of exemplars (Nosofsky 1984), sets of typical features (Tversky 1977), or something else. A different idea is that at least some categories might be defined by shared causal structure: two objects fall under the same concept if and only if they have the same underlying causal structure (Rehder 2003a; b; Rehder and Kim 2006). Specifically, the causal model theory holds that some concepts are defined by a common causal structure, almost always

expressed as a causal Bayes net, and that conceptual reasoning is essentially causal reasoning. For example, causal model theory holds that similarity judgments--how similar some new object  $O$  is to category  $A$ --are given by  $P(O|A)$ : the probability that an object randomly chosen from  $A$  would be like  $O$  (Rehder 2003b), and then those similarities are used to produce categorization judgments (i.e.,  $P(A|O)$ ).<sup>v</sup> Feature inference--given that object  $O$  of type  $A$  has features  $F_1, F_2$ , etc., how likely it is that  $O$  has feature  $G$ --is similarly understood as causal reasoning: probabilistic inference in a particular causal Bayes net given observations of some of the variables (Rehder and Hastie 2004). Causal model theory has led to numerous experimental results that demonstrate clearly the importance of causal beliefs and reasoning in people's concepts, though substantial open questions remain (e.g., its scope of applicability, and its ability to represent conceptual hierarchies).

Causal reasoning is also closely connected to counterfactual reasoning, as our causal knowledge often plays a role in assessing counterfactuals, and counterfactual "but for" reasoning is frequently part of causal reasoning. Psychological research on counterfactual reasoning has focused on both evaluation of the truth of particular counterfactuals, and on the spontaneous generation of counterfactuals (Mandel, Hilton and Catellani 2005). Suppose that factors  $C_1, \dots, C_n$  and outcome  $E$  all occur. People are more likely to judge the counterfactual "If not- $C_i$ , not- $C_j, \dots$ , then not- $E$ " as true to the extent that the factors in the antecedent (a) are anomalous; (b) are controllable; (c) violate a social or moral norm; (d) are close in time or space to the outcome; and/or (e) have a known mechanism connecting them to  $E$ . The same dimensions seem to be relevant for which factors are mentioned in the antecedent of spontaneously generated counterfactuals (Byrne 2005; Roese 1997 provide reviews). One clear conclusion is that, although causal and counterfactual reasoning are closely connected, they are not identical with

one another (Mandel 2003). As an example, consider an individual who takes an unusual route home, but is hit by a drunk driver during the drive. When asked to think about relevant counterfactuals for this individual, most people respond: “if she hadn’t taken the unusual route, then she wouldn’t have been involved in the accident.” At the same time, most people judge the drunk driver to be the principal cause of the accident. More generally, the factors that are weighted most heavily in counterfactual reasoning are not necessarily the ones that are judged to either have the greatest causal influence or be the most causally relevant. Counterfactual and causal reasoning make use of each other, but are not the same cognitive process.

These previous lines of research all provided indirect ways to study causal reasoning. A direct approach is to study people’s causal reasoning when they have to determine the causes of some token event. For example, people presumably decide which event in some sequence caused a car accident by causal reasoning about their prior beliefs. Psychological research on this problem of causal attribution has primarily focused on whether people use knowledge of mechanisms or of correlations to make these decisions (Ahn and Bailenson 1996; Ahn *et al.* 1995). Suppose, for example, that I know that taking some medication is correlated with car accidents, and I know a mechanism by which wet roads lead to car accidents. When asked about the cause of a car accident involving both the medication and a wet road, I am more likely to attribute the accident to the wetness of the road. In general, people prefer to use mechanism information, and when both types of information are used, they weight mechanism information more heavily (Ahn and Bailenson 1996; Ahn *et al.* 1995). This result is not particularly surprising, as knowledge of a mechanism usually implies knowledge of when it is (and is not) likely to be active; knowledge of correlations often does not have the same type of scope knowledge. This dependence on mechanism knowledge in causal attribution is particularly

striking, however, given that most people have an “illusion of explanatory depth” (Rozenblit and Keil 2002): an overestimation of their own mechanism knowledge, and difficulty accepting that their knowledge is limited in this regard. In general, people believe that they can explain the mechanism  $M$  underlying  $X \rightarrow Y$ , and then explain the mechanisms underlying  $M$ , and so on; actually, they can rarely describe anything more than  $M$ . Importantly, however, the notion of ‘mechanism’ used in this research is much broader and weaker than the notion recently advanced in the philosophy of science (e.g., Craver 2007; Machamer, Darden and Craver 2000). The mechanism information in Ahn *et al.*’s studies principally consists of intermediate events or variables, rather than any knowledge of how the pieces fit together (Danks 2005).

## 6. Causality in non-human animals

A final source of information about the psychology of causation comes from comparative studies with other animal species. Historically, animal research has focused on classical and instrumental conditioning (described in section 3). Non-human animals can certainly predict future events, but they were not thought to have any substantive notion of causation in the world; it was assumed that the predictions were based entirely on learned associations, or relatively domain-specific triggers (e.g., Lavin, Freise and Coombes 1980). This consensus view then shifted, particularly with respect to non-human primates, as numerous field studies emerged that reported extensive and seemingly sophisticated tool use by wild animals. For example, chimpanzees were found to “fish” for termites by inserting long, flexible implements (grass, reeds, sticks, etc.) into termite mounds, waiting for the termites to latch onto the implement, and then removing it for a tasty termite snack (Goodall 1986). Chimpanzees were even observed to modify their tools in ways that improved their performance. These observations led many

authors to argue that some non-human animals have a relatively rich notion of causality, and perhaps even the same concept as humans (McGrew 1992; Premack 1976). These claims were principally about non-human primates, though there have periodically been similar claims about other non-human animals.

In recent years, however, the view that non-human animals--primates in particular--have a rich, almost-human notion of causation has come under increased attack. Tomasello and Call (1997) surveyed a wide range of primate behaviors and argued that non-human primates have only very rich associations, rather than any notion of causality that is abstract, or domain-general, or involves unobserved forces. That is, perhaps the remarkably complex behavior exhibited by non-human primates arises from remarkably complex, but entirely perception-based, associations between object shapes, action sequences, and outcomes. Povinelli and colleagues have carried out a number of experiments (summarized in Povinelli 2000) that explicitly try to determine if a chimpanzee's learning involves abstract causal knowledge. Many of their experiments find that chimpanzees are seemingly insensitive to the underlying causal structure of a situation, and respond only to superficial perceptual cues. For example, chimpanzees seem to think that a rope that is touching an object can always be used to pull that object towards them; they do not seem to be sensitive to the fact that a physical *connection* is required, not merely physical contact (Povinelli 2000, ch. 9). If given enough trials, then chimpanzees would of course "learn" the difference between a rope on top of a banana and one tied around the banana, but only through repeated associations between various perceptions (rope on top vs. rope tied around) and success or failure in obtaining the banana. The chimpanzee's eventual success would not be based on reasoning or learning with an abstract, domain-general notion of causation that requires physical connection to manifest. More generally, one can argue

that essentially all findings of “causal learning” or “causal reasoning” in non-human animals can be explained in similar ways (Penn and Povinelli 2007).

These findings have been used to argue for a positive hypothesis about the uniqueness of human behavior (Penn and Povinelli 2007; Penn, Holyoak and Povinelli 2008; Povinelli 2000). The “reinterpretation hypothesis” holds that only humans are able to reinterpret the surface features of the world in terms of complex, necessarily unobserved predicates and relations (e.g., causality, support, force, etc.). In particular, it holds that only humans can take a sequence of perceptual inputs, and explain or describe that sequence in terms of causality. The reinterpretation hypothesis essentially says that Hume’s problem never arises for non-human animals, since they never (re-)conceptualize the world in terms of unobserved causal influences. Of course, non-human animals have concepts of varying complexity, but the reinterpretation hypothesis argues that those concepts are always restricted to the perceptual.

Much of the work on causal learning and reasoning in humans points towards us having a complex, multi-faceted concept, or perhaps even multiple concepts of causation (separate for causal perception and causal inference). Even if we have only a single concept of causation, it surely involves many different dimensions, properties, and relations. We must therefore take care that our understanding of the notion of causation in non-human animals is not based on some simple dichotomy of associationism vs. full-blooded causal learning/reasoning. We should take seriously the possibility that, even if non-human animals do not have the same notion of causality as humans, they need not be “mere” associationists. Two recent experiments with rats illustrate the large middle ground between the endpoints of the standard, simplistic dichotomy.

One standard tenet of associationism is that observations involving a cue (e.g., a tone) affect only the associative strengths of that cue, and perhaps also the strengths of other cues that

have reliably co-occurred with that cue in the past. If a tone has never occurred with a light, for example, then further observations of the tone should not (on the standard account) affect associations involving the light. This assumption turns out to be false, as rats use observations of one cue to revise associative strengths of cues that have never co-occurred with the original cue (Denniston *et al.* 2003).<sup>vi</sup> The rats' behavior suggests that they are using some sort of higher-order "reasoning" about the relationships between the various cues, though the exact nature of that reasoning is currently unknown. At the very least, the "associationist" processes of rats are substantially more sophisticated than the standard account suggests.

An even more striking finding comes from Blaisdell *et al.* (2006), who have recently argued that rats seem to do causal reasoning using (something like) causal Bayes nets. Consider two different causal structures: (a)  $X \leftarrow Y \rightarrow Z$ ; and (b)  $X \rightarrow Y \rightarrow Z$ . Blaisdell *et al.* (2006) split their rats into two groups, and used classical conditioning to "teach" each group both edges in one of the causal structures. For example, group (a) received repeated trials of  $Y$  followed by  $X$ , and separate trials of  $Y$  followed by  $Z$ ; in both groups,  $X$ =tone;  $Y$ =light; and  $Z$ =food. Now consider an intervention to bring about  $X$ . In the common-cause structure (a), the intervention will break the  $X \leftarrow Y$  causal connection, and so neither  $Y$  nor  $Z$  will change; in the chain structure (b), the intervention will not change the causal structure, and so both  $Y$  and  $Z$  will (probabilistically) change. Both groups of rats were provided with such an "intervention" in the form of a lever that produced the tone  $X$ . Rats that "learned" the common-cause structure were significantly less likely to check for the food after pressing the lever, compared to rats that "learned" the chain structure. That is, the rats behaved as if they knew whether their actions to produce  $X$  were likely to bring about a change in  $Z$ , and they acted consistently with the predictions of a causal Bayes net model. This finding does not, of course, prove that rats have

causal Bayes nets “in their heads.” There are more minimal interpretations of this result, and prediction given interventions is only one aspect of causal Bayes nets (Penn and Povinelli 2007). This finding does show, however, that rats are more than just simple associationists: they seem to be able to integrate distinct pieces of observational evidence into a single, relatively coherent structure, and then use that observational evidence to make predictions about the outcomes of interventions (see also Leising *et al.* forthcoming).

Causal learning and reasoning in humans--the full array, scope, and types--seem to be unique among the animal kingdom, and the nature and source of this uniqueness is the subject of ongoing debate (e.g., Penn *et al.* 2008 and accompanying commentaries). Non-human animals are capable of remarkably sophisticated behavior that takes advantage of causal relations in the world to help them reach their goals, but seemingly always in ways that differ crucially, though not always obviously, from human behavior. Research on non-human animals is nonetheless potentially able to provide some insight into the human causal learning and reasoning precisely by revealing the multi-faceted nature of our causal concept(s) and cognition. The ways in which non-human animals are more than simple associationists potentially indicate some of the components of causal cognition in humans. One significant open question is how to use philosophical research on the many dimensions and uses of causation to inform research on which components of human causal cognition are found in non-human animals.

## 7. Conclusion

Psychological research on causation has expanded rapidly in the last twenty years, and it seems to be one of the “hot areas” in cognitive science right now. In these years, there have been significant theoretical and empirical advances on causal perception, inference, and reasoning,

though many open questions remain in all three areas. One striking feature of this research is that it has almost all focused on causal cognition in isolation. For example, an experiment will ask participants to learn some causal structure from observed cases, but then only ask for explicit, verbal causal strength judgments. People are almost never asked to make meaningful decisions using the causal information that they learn. In contrast with these laboratory experiments, causal cognition “in the wild” cannot easily be isolated from other cognitive processes. One of the most significant challenges facing psychologists in coming years is thus to understand better the relationship between causal cognition and other cognitive processes, such as decision-making, linguistic inferences/pragmatics, and social behavior.

## Further Reading

Sloman (2005): Describes many of the ways that causal knowledge and reasoning are relevant to other parts of cognition. Provides an overview of causal Bayes nets and defends them as a psychological account of causal knowledge.

Michotte (1963): Classic text on causal perception. Includes a wide range of experiments on the launching effect.

Scholl and Tremoulet (2000): More recent review article on the launching effect. Argues that causal perception is modular (or close to it).

Gopnik *et al.* (2004): Major paper defending the causal Bayes net view of causal representation and learning. Describes much of the developmental evidence about causal learning.

Danks (2007b): Provides an overview and unification of all of the major theories of causal inference.

Gopnik and Schulz (2007): Edited book covering both causal inference and causal reasoning. Most papers either use, or respond to theories that use, causal Bayes nets.

Sloman and Hagmayer (2006): Presents a theory of decision-making based on causal knowledge, represented as causal Bayes nets.

Mandel (2003): Surveys much of the empirical data about causal and counterfactual reasoning, and argues that they are importantly distinct.

Povinelli (2000): Describes many of the experiments that expose the limits of causal knowledge in non-human primates. Carefully explores methodological challenges facing research on animal cognition.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

Penn and Povinelli (2007): Critical review of research on causal cognition in non-human animals. Questions whether non-human animals have a rich notion of causation.

## References

- Ahn, W.-K. and Bailenson, J. (1996). 'Causal Attribution as a Search for Underlying Mechanisms: An Explanation of the Conjunction Fallacy and the Discounting Principle'. *Cognitive Psychology*, 31: 82-123.
- Ahn, W.-K., Kalish, C. W., Medin, D. L. and Gelman, S. A. (1995). 'The Role of Covariation Versus Mechanism Information in Causal Attribution'. *Cognition*, 54: 299-352.
- Ahn, W.-K., Kim, N. S., Lassaline, M. E. and Dennis, M. J. (2000). 'Causal Status as a Determinant of Feature Centrality'. *Cognitive Psychology*, 41: 361-416.
- Beasley, N. A. (1968). 'The Extent of Individual Differences in the Perception of Causality'. *Canadian Journal of Psychology*, 22: 399-407.
- Blaisdell, A. P., Sawa, K., Leising, K. J. and Waldmann, M. R. (2006). 'Causal Reasoning in Rats'. *Science*, 311: 1020-1022.
- Blakemore, S.-J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., Segebarth, C. and Decety, J. (2001). 'How the Brain Perceives Causality: An Event-Related fMRI Study'. *NeuroReport*, 12: 3741-3746.
- Buehner, M. J. and May, J. (2002). 'Knowledge Mediates the Timeframe of Covariation Assessment in Human Causal Induction'. *Thinking & Reasoning*, 8: 269-295.
- (2003). 'Rethinking Temporal Contiguity and the Judgement of Causality: Effects of Prior Knowledge, Experience, and Reinforcement Procedure'. *The Quarterly Journal of Experimental Psychology*, 56A: 865-890.
- Buehner, M. J., Cheng, P. W. and Clifford, D. (2003). 'From Covariation to Causation: A Test of the Assumption of Causal Power'. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29: 1119-1140.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*.  
Cambridge, MA: The MIT Press.

Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.

Catena, A., Maldonado, A. and Candido, A. (1998). 'The Effect of the Frequency of Judgment and the Type of Trials on Covariation Learning'. *Journal of Experimental Psychology: Human Perception and Performance*, 24: 481-495.

Cheng, P. W. (1997). 'From Covariation to Causation: A Causal Power Theory'. *Psychological Review*, 104: 367-405.

Cheng, P. W. and Novick, L. R. (1990). 'A Probabilistic Contrast Model of Causal Induction'. *Journal of Personality and Social Psychology*, 58: 545-567.

--- (1992). 'Covariation in Natural Causal Induction'. *Psychological Review*, 99: 365-382.

Choi, H. and Scholl, B. J. (2004). 'Effects of Grouping and Attention on the Perception of Causality'. *Perception & Psychophysics*, 66: 926-942.

--- (2006). 'Perceiving Causality after the Fact: Postdiction in the Temporal Dynamics of Causal Perception'. *Perception*, 35: 385-399.

Cohen, L. B. and Amsel, G. N. (1998). 'Precursors to Infants' Perception of the Causality of a Simple Event'. *Infant Behavior and Development*, 21: 713-731.

Collins, D. J. and Shanks, D. R. (2006). 'Conformity to the Power PC Theory of Causal Induction Depends on the Type of Probe Question'. *The Quarterly Journal of Experimental Psychology*, 59: 225-232.

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., Robbins, T. W., Bullmore, E. T. and Fletcher, P. C. (2004). 'Prediction Error During

Retrospective Reevaluation of Causal Associations in Humans: fMRI Evidence in Favor of an Associative Model of Learning'. *Neuron*, 44: 877-888.

Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Csibra, G., Gergely, G., Bíró, S., Koós, O. and Brockbank, M. (1999). 'Goal Attribution without Agency Cues: The Perception of 'Pure Reason' in Infancy'. *Cognition*, 72: 237-267.

Danks, D. (2003). 'Equilibria of the Rescorla-Wagner Model'. *Journal of Mathematical Psychology*, 47: 109-121.

--- (2005). 'The Supposed Competition between Theories of Human Causal Inference'. *Philosophical Psychology*, 18: 259-272.

--- (2007a). 'Theory Unification and Graphical Models in Human Categorization', in A. Gopnik and L. E. Schulz (eds.), *Causal Learning: Psychology, Philosophy, and Computation*. Oxford: Oxford University Press, 173-189.

--- (2007b). 'Causal Learning from Observations and Manipulations', in M. C. Lovett and P. Shah (eds.), *Thinking with Data*. Mahwah, NJ: Lawrence Erlbaum Associates, 359-388.

Danks, D., Griffiths, T. L. and Tenenbaum, J. B. (2003). 'Dynamical Causal Learning', in S. Becker, S. Thrun and K. Obermayer (eds.), *Advances in Neural Information Processing Systems 15*. Cambridge, MA: The MIT Press, 67-74.

De Houwer, J. and Beckers, T. (2002). 'A Review of Recent Developments in Research and Theories on Human Contingency Learning'. *The Quarterly Journal of Experimental Psychology*, 55B: 289-310.

Denniston, J. C., Savastano, H. I., Blaisdell, A. P. and Miller, R. R. (2003). 'Cue Competition as a Retrieval Deficit'. *Learning and Motivation*, 34: 1-31.

- In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.
- Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R. A. E., Carpenter, T. A., Donovan, T., Papadakis, N. and Bullmore, E. T. (2001). 'Responses of Human Frontal Cortex to Surprising Events Are Predicted by Formal Associative Learning Theory'. *Nature Neuroscience*, 4: 1043-1048.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: The MIT Press.
- Fonlupt, P. (2003). 'Perception and Judgement of Physical Causality Involve Different Brain Structures'. *Cognitive Brain Research*, 17: 248-254.
- Fugelsang, J. A. and Thompson, V. A. (2003). 'A Dual-Process Model of Belief and Evidence Interactions in Causal Reasoning'. *Memory & Cognition*, 31: 800-815.
- Fugelsang, J. A. and Dunbar, K. N. (2005). 'Brain-Based Mechanisms Underlying Complex Causal Thinking'. *Neuropsychologia*, 42: 1204-1213.
- Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S. and Dunbar, K. N. (2005). 'Brain Mechanisms Underlying Perceptual Causality'. *Cognitive Brain Research*, 24.
- Gergely, G., Nádasdy, Z., Csibra, G. and Bíró, S. (1995). 'Taking the Intentional Stance at 12 Months of Age'. *Cognition*, 56: 165-193.
- Goldvarg, E. and Johnson-Laird, P. N. (2001). 'Naive Causality: A Mental Model Theory of Causal Meaning and Reasoning'. *Cognitive Science*, 25: 565-610.
- Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, MA: Harvard University Press.
- Gopnik, A. and Schulz, L. E. (eds.) (2007). *Causal Learning: Psychology, Philosophy, and Computation*, Oxford: Oxford University Press.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

- Gopnik, A. and Glymour, C. (2002). 'Causal Maps and Bayes Nets: A Cognition and Computational Account of Theory-Formation', in P. Carruthers, S. Stich and M. Siegal (eds.), *The Cognitive Basis of Science*. Cambridge: Cambridge University Press, 117-132.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T. and Danks, D. (2004). 'A Theory of Causal Learning in Children: Causal Maps and Bayes Nets'. *Psychological Review*, 111: 3-32.
- Griffiths, T. L. and Tenenbaum, J. B. (2005). 'Structure and Strength in Causal Induction'. *Cognitive Psychology*, 51: 334-384.
- Hagmayer, Y. and Waldmann, M. R. (2002). 'How Temporal Assumptions Influence Causal Judgments'. *Memory & Cognition*, 30: 1128-1137.
- Hagmayer, Y. and Sloman, S. A. (2005). 'A Causal Model Theory of Choice', in B. G. Bara, L. Barsalou and M. Bucciarelli (eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum, 881-886.
- Hagmayer, Y., Sloman, S. A., Lagnado, D. A. and Waldmann, M. R. (2007). 'Causal Reasoning through Intervention', in A. Gopnik and L. E. Schulz (eds.), *Causal Learning: Psychology, Philosophy, and Computation*. Oxford: Oxford University Press, 86-100.
- Heider, F. and Simmel, M.-A. (1944). 'An Experimental Study of Apparent Behavior'. *American Journal of Psychology*, 57: 243-249.
- Lagnado, D. A. and Sloman, S. A. (2004). 'The Advantage of Timely Intervention'. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30: 856-876.
- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. and Sloman, S. A. (2007). 'Beyond Covariation: Cues to Causal Structure', in A. Gopnik and L. E. Schulz (eds.), *Causal*

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

*Learning: Psychology, Philosophy, and Computation*. Oxford: Oxford University Press, 154-172.

Lavin, M. J., Freise, B. and Coombes, S. (1980). 'Transferred Flavor Aversions in Adult Rats'. *Behavioral and Neural Biology*, 28: 15-33.

Leising, K. J., Wong, J., Waldmann, M. R. and Blaisdell, A. P. (forthcoming). 'The Special Status of Actions in Causal Reasoning in Rats'. *Journal of Experimental Psychology: General*.

Leslie, A. M. (1982). 'The Perception of Causality in Infants'. *Perception*, 11: 173-186.

--- (1984). 'Spatiotemporal Continuity and the Perception of Causality in Infants'. *Perception*, 13: 287-305.

--- (1994). 'ToMM, ToBy, and Agency: Core Architecture and Domain Specificity', in L. Hirschfield and S. A. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: Cambridge University Press, 119-148.

Leslie, A. M. and Keeble, S. (1987). 'Do Six-Month-Old Infants Perceive Causality?' *Cognition*, 25: 265-288.

Lober, K. and Shanks, D. R. (2000). 'Is Causal Induction Based on Causal Power? Critique of Cheng (1997)'. *Psychological Review*, 107: 195-212.

Machamer, P., Darden, L. and Craver, C. F. (2000). 'Thinking About Mechanisms'. *Philosophy of Science*, 67: 1-25.

Mandel, D. R. (2003). 'Judgment Dissociation Theory: An Analysis of Differences in Causal, Counterfactual, and Covariational Reasoning'. *Journal of Experimental Psychology: General*, 132: 419-434.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

Mandel, D. R., Hilton, D. J. and Catellani, P. (eds.) (2005). *The Psychology of Counterfactual Thinking*, New York: Routledge.

McGrew, W. C. (1992). *Chimpanzee Material Culture: Implications for Human Evolution*. Cambridge: Cambridge University Press.

McKenzie, C. R. M. and Mikkelsen, L. A. (2007). 'A Bayesian View of Covariation Assessment'. *Cognitive Psychology*, 54: 33-61.

Michotte, A. (1963). *The Perception of Causality*. London: Methuen.

Morris, M. W. and Peng, K. (1994). 'Culture and Cause: American and Chinese Attributions for Social and Physical Events'. *Journal of Personality and Social Psychology*, 67: 949-971.

Newman, G. E., Choi, H., Wynn, K. and Scholl, B. J. (forthcoming). 'The Origins of Causal Perception: Evidence from Postdictive Processing in Infancy'. *Cognitive Psychology*.

Nichols, W. and Danks, D. (2007). 'Decision Making Using Learned Causal Structures', in D. S. McNamara and J. G. Trafton (eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 1343-1348.

Nosofsky, R. M. (1984). 'Choice, Similarity, and the Context Theory of Classification'. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10: 104-114.

Novick, L. R. and Cheng, P. W. (2004). 'Assessing Interactive Causal Influence'. *Psychological Review*, 111: 455-485.

Oakes, L. M. (1994). 'Development of Infants' Use of Continuity Cues in Their Perception of Causality'. *Developmental Psychology*, 30: 869-879.

Oakes, L. M. and Cohen, L. B. (1990). 'Infant Perception of a Causal Event'. *Cognitive Development*, 5: 193-207.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

Pearce, J. M. (1987). 'A Model for Stimulus Generalization in Pavlovian Conditioning'.

*Psychological Review*, 94: 61-73.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge

University Press.

Penn, D. C. and Povinelli, D. J. (2007). 'Causal Cognition in Humans and Nonhuman Animals:

A Comparative, Critical Review'. *Annual Review of Psychology*, 58: 97-118.

Penn, D. C., Holyoak, K. J. and Povinelli, D. J. (2008). 'Darwin's Mistake: Explaining the

Discontinuity between Human and Nonhuman Minds'. *Behavioral and Brain Sciences*,

31: 109-130.

Perales, J. C. and Shanks, D. R. (2007). 'Models of Covariation-Based Causal Judgment: A

Review and Synthesis'. *Psychonomic Bulletin & Review*, 14: 577-596.

Posner, M. I. and Keele, S. W. (1968). 'On the Genesis of Abstract Ideas'. *Journal of*

*Experimental Psychology*, 77: 353-363.

Povinelli, D. J. (2000). *Folk Physics for Apes: The Chimpanzee's Theory of How the World*

*Works*. Oxford: Oxford University Press.

Premack, D. (1976). *Intelligence in Ape and Man*. Hillsdale, NJ: Erlbaum.

Rehder, B. (2003a). 'A Causal-Model Theory of Conceptual Representation and Categorization'.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29: 1141-1159.

--- (2003b). 'Categorization as Causal Reasoning'. *Cognitive Science*, 27: 709-748.

Rehder, B. and Hastie, R. (2004). 'Category Coherence and Category-Based Property Induction'.

*Cognition*, 91: 113-153.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

Rehder, B. and Kim, S. (2006). 'How Causal Knowledge Affects Classification: A Generative Theory of Categorization'. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 32: 659-683.

Rescorla, R. A. and Wagner, A. R. (1972). 'A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement', in A. H. Black and W. F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory*. New York: Appleton-Century-Crofts, 64-99.

Roese, N. J. (1997). 'Counterfactual Thinking'. *Psychological Bulletin*, 121: 133-148.

Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M. and Gazzaniga, M. S. (2005). 'Dissociating Processes Supporting Causal Perception and Causal Inference in the Brain'. *Neuropsychology*, 19: 591-602.

Rozenblit, L. and Keil, F. C. (2002). 'The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth'. *Cognitive Science*, 26: 521-562.

Saffran, J. R., Aslin, R. N. and Newport, E. L. (1996). 'Statistical Learning by 8-Month-Old Infants'. *Science*, 274: 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N. and Newport, E. L. (1999). 'Statistical Learning of Tone Sequences by Human Infants and Adults'. *Cognition*, 70: 27-52.

Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J. and Lieberman, M. D. (2005). 'An fMRI Study of Causal Judgments'. *European Journal of Neuroscience*, 22: 1233-1238.

Schlottmann, A. (1999). 'Seeing It Happen and Knowing How It Works: How Children Understand the Relation between Perceptual Causality and Underlying Mechanism'. *Developmental Psychology*, 35: 303-317.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

--- (2000). 'Is Perception of Causality Modular?' *Trends in Cognitive Sciences*, 4: 441-442.

Schlottmann, A. and Shanks, D. R. (1992). 'Evidence for a Distinction between Judged and Perceived Causality'. *Quarterly Journal of Experimental Psychology*, 44A: 321-342.

Schlottmann, A. and Anderson, N. H. (1993). 'An Information Integration Approach to Phenomenal Causality'. *Memory & Cognition*, 21: 785-801.

Schlottmann, A., Ray, E. D., Mitchell, A. and Demetriou, N. (2006). 'Perceived Physical and Social Causality in Animated Motions: Spontaneous Reports and Ratings'. *Acta Psychologica*, 123: 112-143.

Scholl, B. J. and Tremoulet, P. D. (2000). 'Perceptual Causality and Animacy'. *Trends in Cognitive Sciences*, 4: 299-309.

Scholl, B. J. and Nakayama, K. (2002). 'Causal Capture: Contextual Effects on the Perception of Collision Events'. *Psychological Science*, 13: 493-498.

Schulz, L. E. and Gopnik, A. (2004). 'Causal Learning across Domains'. *Developmental Psychology*, 40: 162-176.

Shanks, D. R. (1995). 'Is Human Learning Rational?' *The Quarterly Journal of Experimental Psychology*, 48A: 257-279.

Sloman, S. A. (2005). *Causal Models: How People Think About the World and Its Alternatives*. Oxford: Oxford University Press.

Sloman, S. A. and Hagmayer, Y. (2006). 'The Causal Psycho-Logic of Choice'. *Trends in Cognitive Sciences*, 10: 407-412.

Sobel, D. M. and Kushnir, T. (2006). 'The Importance of Decision Making in Causal Learning from Interventions'. *Memory & Cognition*, 34: 411-419.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction, and Search*. Berlin: Springer-Verlag.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J. and Blum, B. (2003). 'Inferring Causal Networks from Observations and Interventions'. *Cognitive Science*, 27: 453-489.

Tenenbaum, J. B. and Griffiths, T. L. (2001). 'Structure Learning in Human Causal Induction', in T. Leen, T. Deitterich and V. Tresp (eds.), *Advances in Neural Information Processing Systems 13*. Cambridge, MA: The MIT Press, 59-65.

Tomasello, M. and Call, J. (1997). *Primate Cognition*. Oxford: Oxford University Press.

Turner, D. C., Aitken, M. R. F., Shanks, D. R., Sahakian, B. J., Robbins, T. W., Schwarzbauer, C. and Fletcher, P. C. (2004). 'The Role of the Lateral Frontal Cortex in Causal Associative Learning: Exploring Preventative and Super-Learning'. *Cerebral Cortex*, 14: 872-880.

Tversky, A. (1977). 'Features of Similarity'. *Psychological Review*, 84: 327-352.

Uleman, J. S., Saribay, S. A. and Gonzalez, C. M. (2008). 'Spontaneous Inferences, Implicit Impressions, and Implicit Theories'. *Annual Review of Psychology*, 59: 329-360.

Van Hamme, L. J. and Wasserman, E. A. (1994). 'Cue Competition in Causality Judgments: The Role of Nonpresentation of Compound Stimulus Elements'. *Learning and Motivation*, 25: 127-151.

Waldmann, M. R. (1996). 'Knowledge-Based Causal Induction', in D. R. Shanks, K. J. Holyoak and D. L. Medin (eds.), *Causal Learning: The Psychology of Learning and Motivation, Vol. 34*. San Diego, CA: Academic Press, 47-88.

Waldmann, M. R. and Hagmayer, Y. (2006). 'Categories and Causality: The Neglected Direction'. *Cognitive Psychology*, 53 27-58.

In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation*. Oxford: Oxford University Press.

White, P. A. (1995). *The Understanding of Causation and the Production of Action*. Hillsdale, NJ: Erlbaum.

--- (2003a). 'Causal Judgement as Evaluation of Evidence: The Use of Confirmatory and Disconfirmatory Information'. *The Quarterly Journal of Experimental Psychology*, 56A: 491-513.

--- (2003b). 'Effects of Wording and Stimulus Format on the Use of Contingency Information in Causal Judgment'. *Memory & Cognition*, 31: 231-242.

--- (2003c). 'Making Causal Judgments from the Proportion of Confirming Instances: The pCI Rule'. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29: 710-727.

White, P. A. and Milne, A. (1997). 'Phenomenal Causality: Impressions of Pulling in the Visual Perception of Objects in Motion'. *American Journal of Psychology*, 110: 573-602.

--- (1999). 'Impressions of Enforced Disintegration and Bursting in the Visual Perception of Collision Events'. *Journal of Experimental Psychology: General*, 128: 499-516.

Wolff, P. (2007). 'Representing Causation'. *Journal of Experimental Psychology: General*, 136: 82-111.

## Endnotes

---

<sup>i</sup> Two omissions merit special mention. First, there is an enormous literature in social psychology on causal attributions as one type of social inference (see Uleman, Saribay and Gonzalez 2008 for a recent review). Unfortunately, space considerations prevent any serious examination of that research, though social inference is briefly discussed in section 2. Second, there is a small-but-growing body of experimental research on people's explicit (but untutored) judgments about the "meaning" of the word 'cause' (e.g., Goldvarg and Johnson-Laird 2001; Wolff 2007). This research is still very much in its infancy, and there is growing evidence that the word 'cause' is linguistically ambiguous.

<sup>ii</sup> Many examples of such sequences can be found at the website of Brian Scholl's research group: <http://www.yale.edu/perception/Brian/demos/causality.html>

<sup>iii</sup> One might wonder how we could determine such a thing, given that six-month-olds are pre-verbal, and even pre-mobile. All of the cited experiments are so-called "looking time studies." The basic idea underlying this experimental design is that infants look longer at things that interest them, and stop looking at things that bore them. Thus, if infants who have repeatedly seen  $Q$  subsequently look longer at  $A$  rather than  $B$ , then those infants must think that  $A$  is more different from  $Q$  than  $B$  is. If  $Q$ ,  $A$ , and  $B$  are appropriately matched on all-but-one dimensions, then the infants are arguably conceptualizing (or perceiving)  $Q$  and  $B$  as the same on that last dimension, while  $A$  is different. There are obvious concerns about looking time studies, and they are notoriously difficult to interpret. Nonetheless, this experimental method is the best we currently have for understanding the mental life of infants.

<sup>iv</sup> This operation severs the corpus callosum: the (large) neural connection between the two hemispheres of the brain. It is most commonly performed as a "last-resort" attempt to control

seizures. Cognitive processes can sometimes be localized in these patients to one hemisphere or the other by presenting information to only one eye at a time.

<sup>v</sup> Early statements of causal model theory did not draw a clean distinction between similarity and categorization judgments-- $P(O|A)$  and  $P(A|O)$ , respectively--largely because the experimental designs often did not separate the two. More recent statements of causal model theory seem to have converged on the version presented here (Danks 2007a).

<sup>vi</sup> Specifically, the rats were shown a series of  $AX+$  trials (i.e., cue  $A$  with cue  $X$ , followed by the outcome), then a series of  $XY+$  trials. At this point, the rats have a relatively strong association between  $Y$  and the outcome. A subsequent series of  $A-$  trials (i.e., just cue  $A$ , and no outcome) led to a reduced associative strength for  $Y$ . The rats' use of the  $A-$  trials to retrospectively increase the associative strength of  $X$  (to "explain" the  $AX+$  trials) is not novel. The surprising part is that the rats seemingly propagate that change outward (in some sense) to retrospectively revise their learning from the  $XY+$  trials: increased strength for  $X$  means less learned strength for  $Y$ .