



Causality in Thought

Steven A. Sloman¹ and David Lagnado²

¹Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island 02912; email: Steven_Sloman@brown.edu

²Cognitive, Perceptual, and Brain Sciences Department, University College London, WC1E 6BT United Kingdom; email: d.lagnado@ucl.ac.uk

Annu. Rev. Psychol. 2015. 66:3.1–3.25

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi:
10.1146/annurev-psych-010814-015135

Copyright © 2015 by Annual Reviews.
All rights reserved

Keywords

causal reasoning, causal judgment, causal decision making, causal attribution

Abstract

Causal knowledge plays a crucial role in human thought, but the nature of causal representation and inference remains a puzzle. Can human causal inference be captured by relations of probabilistic dependency, or does it draw on richer forms of representation? This article explores this question by reviewing research in reasoning, decision making, various forms of judgment, and attribution. We endorse causal Bayesian networks as the best normative framework and as a productive guide to theory building. However, it is incomplete as an account of causal thinking. On the basis of a range of experimental work, we identify three hallmarks of causal reasoning—the role of mechanism, narrative, and mental simulation—all of which go beyond mere probabilistic knowledge. We propose that the hallmarks are closely related. Mental simulations are representations over time of mechanisms. When multiple actors are involved, these simulations are aggregated into narratives.

Contents

INTRODUCTION.....	3.2
CONDITIONAL REASONING.....	3.4
INTERVENTION AND COUNTERFACTUAL CONDITIONALS.....	3.6
DECISION MAKING AND ATTRIBUTION.....	3.9
Deciding the Actual Cause.....	3.10
Legal Decision Making.....	3.12
Attributing Responsibility.....	3.13
Moral Reasoning.....	3.14
INTERIM CONCLUSION.....	3.14
ASYMMETRY OF CAUSE AND EFFECT.....	3.14
MARKOV VIOLATIONS.....	3.16
ILLUSION OF EXPLANATORY DEPTH.....	3.17
DISCUSSION.....	3.18
On Mechanism.....	3.18
CONCLUSION.....	3.19

INTRODUCTION

Despite what some bankers and poets may think, it is neither love nor money that makes the world go 'round. It is causation. This is not intended as a metaphysical claim, and so we hope not to offend any readers who know some philosophy or who, like Russell (1913), do not believe that scientific laws are causal. Our claim is psychological; we claim only that the human cognitive system is built to see causation as governing how events unfold.

Such a claim opens a minefield. Perhaps the cognitive system doesn't "see" at all; maybe it just acts. Or perhaps it represents events but not what governs them. Nevertheless, our claim has many antecedents. Much of psychology's inspiration for treating causal notions as the infrastructure of thought comes from Michotte (1963), who initiated the attempt to identify how people detect causality in perception. White (2013) has opened new avenues in this tradition by proposing a set of heuristics that the perceptual system applies to identify causal relations. Buehner & Humphreys (2009, 2010) have found that causal relations can actually warp our perception of events and of the time delays between events. Bechlivanidis & Lagnado (2013) show that causal beliefs (even if acquired in a short training period) can make people reorder events to fit with the time course expected by their causal models.

The focus of this review, however, is on thought, not perception, where a number of voices have recently been arguing that causality is central. Early proponents were Waldmann & Holyoak (1992) and Cheng (1997), who made strong cases that judgment is finely tuned to beliefs about causal structure and strength. Even before them, Kahneman & Tversky (1982a) showed that people substitute causal inferences when asked for other sorts of judgments. The floodgates for research were opened by Pearl (2000), who, building on work by Spirtes et al. (1993/2000), developed a normative account of causal representation and inference using an extension of the Bayesian networks formalism commonly called causal Bayes nets.¹ This framework has led to an explosion

¹We include definitions (see Glossary sidebar) to help the reader digest our more technical language.

GLOSSARY

Causal Bayesian networks: a causal model is made up of two components: (a) a graph representing the variables in the model with directed links between variables representing direct causal relations and (b) a set of conditional probability tables (or structural equations) specifying how each variable depends on its direct causes. A causal model supports prediction, inference given interventions on the system, and counterfactual queries.

Do operator: a special operator for representing interventions. An idealized intervention on a variable in a causal system involves (a) setting that variable to a specified value and (b) removing links to it from all the typical causes of that variable. This captures the difference between setting a variable (by intervention) and observing that the variable takes that value. The do operator dictates that when one intervenes on a variable, no diagnostic inferences are drawn about the usual causes of that variable. See **Figure 1** for an illustration.

Pruning theory: as well as supporting inferences given interventions, causal models also support counterfactual inferences. The key idea is that a counterfactual represents an intervention on a mental causal model. Pruning theory represents the claim that counterfactual inference involves applying the do operator to a causal model. For example, to represent the counterfactual claim, “If *X* had not happened, then *Y* would not have happened,” one first updates the causal model according to what actually happened (i.e., both *X* and *Y* actually occurred), then one intervenes on the model by setting *X* to false and removing causal links into *X*; finally, one updates the values of the model given this intervention. This process dictates that there is no backtracking, i.e., that an intervention on *X* does not affect the typical causes of *X*. See **Figure 2** for an illustration.

Counterfactual backtracking: counterfactual backtracking occurs when pruning theory is violated, when a counterfactual assumption is treated as diagnostic of its causes in violation of the do operation.

Minimal network theory: an alternative approach to counterfactual inference that does not assume that the do operator is used. In contrast to pruning theory, minimal network theory assumes that causal principles are not disrupted by counterfactual interventions. When assessing the counterfactual claim “if *X* had not happened, then *Y* would not have happened,” the variable *X* is updated to false, but the causal links into it are not removed. Then all other variables are updated to be made consistent with this change. Hence, backtracking inferences are permitted. See **Figure 2** for an illustration.

Overdetermination: an outcome is overdetermined when there is more than one cause that is sufficient for the outcome. The classic example is when two marksmen shoot a prisoner, and each of their shots is individually sufficient for the death. This case presents a problem for standard counterfactual theories because neither marksman’s shot is necessary for the death: If one marksman hadn’t shot, the prisoner would still have died. See **Figure 3a** for an illustration.

Preemption: preemption occurs when an outcome is caused by one event but would have been caused by another event if the first event had not occurred. For example, marksman *A* shoots first and kills the prisoner, but marksman *B*’s shot would have killed the prisoner if marksman *A* had missed. This case also presents a problem for standard counterfactual theories, because marksman *A*’s shot is not necessary for the death: if he hadn’t shot, the prisoner would still have died. See **Figure 3b** for an illustration.

Double prevention: double prevention occurs when an outcome is caused by one event but there is an attempted prevention that is itself prevented by a third cause. The classic example is when an assassin shoots at his victim, the bodyguard runs to protect the victim, but his attempt to prevent the death is thwarted by someone else who trips the bodyguard. This case is depicted in **Figure 3c**.

of work in psychology and philosophy as well as in technical disciplines; some of this work is summarized by Sloman (2005). Since then, causation has become an even more central topic in the study of thinking. Chater & Oaksford (2013; see also Holyoak & Cheng 2011) argue that causal notions are the key players in human learning, reasoning, perception, and some aspects of language.

In the broadest strokes, the key idea of causal Bayes nets is that we can learn and determine how much certainty we should have in our beliefs by doing Bayesian inference over distributions represented by causal strength and causal structure, along with a special operator for representing intervention (the “do” operator). Should we think about human causal reasoning this way? Should we think about our knowledge of causal relations as a representation of probabilistic dependencies among properties and events? It may be that this view leaves out some important aspects of causal cognition. In particular, the Bayesian view focuses exclusively on probabilistic relations among causes and effects and fails to consider other aspects of the mechanisms by which causes lead to effects. As a result, it may underestimate human reasoning capabilities by failing to represent people’s ability to use spatiotemporal information and other aspects of mechanisms, narratives, and explanations (Lagnado 2011a). Indeed Waldmann & Hagmayer (2013) argue that people represent more about mechanism than the Bayesian covariation perspective allows. At the same time, it may overestimate people’s access to distributional knowledge and ability to compute with it.

This question serves as the organizing principle of this review. Are causal representations primarily built out of probabilistic dependency relations, or are such relations insufficient to explain human thought and behavior? A distinction has been made between two views people can take toward categorization, probability, and inference (Barbey & Sloman 2007, Hampton 2012, Kahneman & Tversky 1982b, Lagnado & Sloman 2004a, Sirota et al. 2014). People can take an outside view by constructing representations out of sets of instances. Or people can take an inside view by constructing representations out of properties of a typical instance. The distinction also applies to models of causal reasoning. Some causal model theories take an outside view and assume that a causal relation is inferred from the relative frequency of possibilities (Cheng 1997, Griffiths & Tenenbaum 2009, Novick & Cheng 2004, Waldmann & Holyoak 1992). Others take an inside view and focus on the structure of single prototypical events. Such theorists think about a causal relation as supporting an intervention (e.g., Waldmann & Hagmayer 2005, Woodward 2003), a counterfactual (Wells & Gavanski 1989), or a force of some kind (Wolff 2007). These views are not mutually exclusive because they refer only to different perspectives on the same object, a causal relation. But, as we discuss below, they represent different kinds of information and support different kinds of inferences.

We start our discussion by revealing why causal notions are so important to understand how people think. Then we explore in greater depth how people think about the core properties of causal relations.

CONDITIONAL REASONING

Reasoning involves inference, deriving new beliefs from information that is known with some degree of certainty or just assumed to be true (for the sake of argument). Causal reasoning involves inference about causes and effects. A causal relation is not merely an association in the sense that it is not a representation of a mere correlation but rather a representation of something more enduring in nature: Natural laws dictate what causes what regardless of any learning. On most views, causal relations are distinct in that they asymmetrically support intervention: Sufficient changes to a cause will (all else being equal) also change the effect; the converse is not true. So causal learning is not merely associative conditioning. Moreover, causation follows a logic (for

instance, the presence of an event raises the probability of the causes of the event), but not any of the standard symbolic logics such as propositional calculus or predicate calculus. To see this, just note that causal logic does not support a standard inference rule of logic, *modus ponens*:

If A then B

A

Therefore, B .

In the case in which A is a cause (e.g., Alan smokes) and B is an effect (Alan's health is poor), and we're following the rules of causal inference, the conclusion (B) does not necessarily follow even if the premises (If A then B , A) are true. Alan could be one of those people who are in great health despite smoking. Propositional logic makes a number of assumptions (like "If A then B " represents the material conditional) that causal logic does not make, and causal logic makes its own assumptions. In general, causal logic can be distinguished from propositional logic in at least three ways: (a) Causal relations are not just stipulated but rather represent mechanisms in the world that take input (causes, enablers, disablers, preventers) and generate outputs (effects). (b) These causal relations unfold over time such that effects cannot precede their causes. (c) If a variable is intervened on, acted on by an agent external to the causal system, then the downstream effects of that variable change but its upstream causes do not.

People find causal reasoning more natural than other forms of reasoning for problems of the same complexity. Perhaps the clearest example of this comes from the work of Cummins (1995; Cummins et al. 1991) in which people evaluated conditional syllogisms such as *modus ponens* (MP), *modus tollens* (MT), *denying the antecedent* (DA), and *affirming the consequent* (AC) with causal content. Consider the following two MP arguments:

(A) If Mary jumped into the swimming pool, then she got wet.

Mary jumped into the swimming pool.

Therefore, she got wet.

(B) If fertilizer was put on the plants, then they grew quickly.

Fertilizer was put on the plants.

Therefore, they grew quickly.

Note that the conditional in (A) has fewer disablers than that in (B): It's harder to imagine how someone would not get wet jumping into a pool than to imagine how a plant might not grow quickly despite having fertilizer put on it. Hence, participants judged (A) more acceptable than (B). This suggests that people did not simply apply a logical schema to evaluate the argument, they also thought about the causal mechanism at stake and the conditions required for it to operate even though they were not asked to do so. The authors make a parallel argument for AC except that instead of considering the availability of disablers, they consider the availability of alternative causes:

(C) If the trigger was pulled, then the gun fired.

The gun fired.

Therefore, the trigger was pulled.

(D) If Mary jumped into the swimming pool, then she got wet.

Mary got wet.

Therefore, she had jumped into the swimming pool.

There are more alternative causes for getting wet than for a gun firing. So the conclusion seems less likely in the case of (D), and people indeed judged it less acceptable. Using some clever paradigms, Cummins and coworkers verified these intuitions: The number of disablers

influenced the acceptability of MP inferences, and the number of alternative causes influenced the acceptability of AC inferences. Her data suggest that people reason about aspects of causation even when they are not asked to do so. Replications and extensions can be found in Chandon & Janiszewski (2009), Markovits & Potvin (2001), and Thompson (1995). Fernbach & Erb (2013) use both a model-fitting strategy to show that causal model theory fit data on MP and AC inferences better than mental model theory (Goldvarg & Johnson-Laird 2001) and a direct experimental test to show that people's inference patterns are indeed causal and that they are not merely computing conditional probabilities (cf. Weidenfeld et al. 2005).

Ali et al. (2011) compared causal model and mental model theories of conditional reasoning by having participants perform diagnostic and predictive causal reasoning tasks. They focused on a phenomenon called “explaining away” (Pearl 1988), also known as “discounting” (Kelley 1973, Morris & Larrick 1995). Explaining away can occur when an effect has more than one independent cause. If the effect occurs, the probability of each of the causes increases (because the effect is diagnostic of the causes). But if one of the causes is known to have occurred as well, then the probabilities of the others go down, potentially down to the level they were at before anything was learned. This is because one explanation for the effect is sufficient; one cause can account for it, so the other causes provide no additional explanatory value and are no more likely. Explaining away is a fundamental assumption of the causal model framework. Mental model theory does not predict it, at least not in the cases that Ali et al. (2011) studied. People's sensitivity to explaining away provides evidence for the causal model framework and against mental model theory.

INTERVENTION AND COUNTERFACTUAL CONDITIONALS

Interventions are changes to a causal system brought about by a variable external to the system. These variables are usually agents assumed to be intentional who have their own motives for intervening. But that is not critical. What is critical is that the variable's action is sufficient to set the value of an endogenous variable and that the action is known to affect the system only through the variable acted on, that there are no backdoor paths to the effects of the action. In that case, beliefs about the normal causes of the variable intervened on should not be changed by the intervention. The change cannot be diagnostic of the endogenous variable's causes because the change is determined by the external agent, not by the normal causes. Intervention is in fact a special case of explaining away (explaining the effect by the action, not by any other variable in the system). For instance, I can intervene on my spouse preparing dessert by turning off the oven while the cake is baking. If the intervention is sufficiently strong, then intervening on a cause changes its effects; intervening on an effect does not change its causes. Turning off the oven will ruin the cake; it will not undo all the hard work that my spouse has already done.

A simple causal Bayes net of this example is shown in **Figure 1**. Nodes in the graph represent variables of interest, and arrows represent direct causal relations. The model states that baking the cake depends on preparing it, that enjoying it depends on baking it, and that there is no direct cause from preparing to enjoying it. An intervention on the baking (say by turning the oven off) will ruin the cake and thus our enjoyment of it, but it will not affect the preparation of the cake. Inference to preparation would be a backtracking inference and is disallowed on the standard interpretation of interventions (Pearl 2000). This conception of intervention is captured formally by invoking the do operator. It sets the baking variable to false and removes incoming causal links to it. The baking failed because of our intervention, not because of poor preparation.

A number of studies have analyzed whether people are sensitive to this asymmetry in the inferential power of intervention by asking people to imagine possible interventions. In one study, Sloman & Lagnado (2005) asked people to imagine a rocket ship with two components, *A* and *B*:

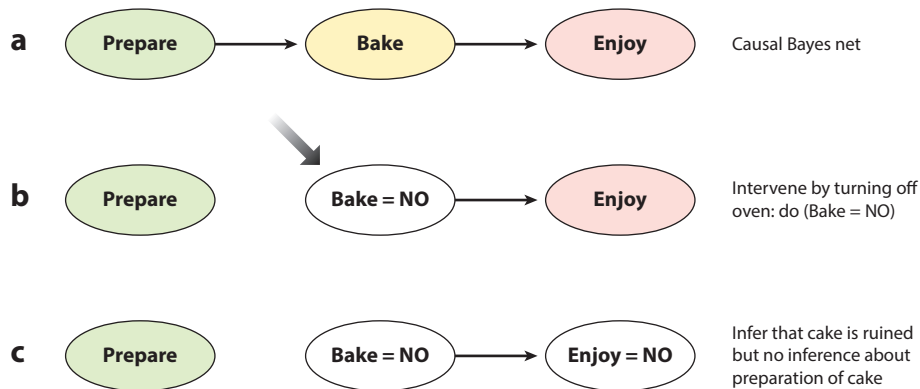


Figure 1

(a) Simple causal Bayes net representing the causal relations between preparation, baking, and enjoyment of a cake. (b) An intervention to stop the baking process. By setting the baking to fail, it removes any causal links to that variable and sets it to a fixed value. (c) Given this intervention, one infers that the cake will not be enjoyed, but one does not infer that the preparation failed.

Movement of component *A* causes component *B* to move. In other words, if *A*, then *B*. Both are moving.

- (a) Suppose component *A* were prevented from moving. Would component *B* still be moving? (Yes or no)
- (b) Suppose component *B* were prevented from moving. Would component *A* still be moving? (Yes or no)

We would expect people to say “no” for (a) because it is in the causal direction. The critical question is (b): Do people backtrack from *B* to make an inference about *A*? Do they treat *B* as diagnostic of *A* and say no? Alternatively, do they apply the do operator, refrain from diagnostic inference, and say yes? A different set of people was asked about observation rather than intervention. They were given the identical information except instead of being asked to suppose *A* were prevented from moving, they were asked to suppose *A* were observed to not be moving.

Observation is different than intervention. Observing an effect is diagnostic of its causes, so people should say no to both observation questions. And that’s what happened. Roughly 80% of people consistently said no when the question concerned observation, and roughly 80% said yes to question (b) when it concerned prevention, a form of intervention. Waldmann & Hagmayer (2005) also found that people distinguish intervention from observation after being taught both chain and common cause structures. These data provide some support for the idea that counterfactual intervention is well represented by the do operator, an idea that lies at the heart of Pearl’s (2000) theory of counterfactual inference (for a recent exposition, see also Pearl 2013).

Rips (2010) and Rips & Edwards (2013) challenge the interventional account. They compare Pearl’s theory (they call it pruning theory) to an alternative suggested by Hiddleston (2005), minimal network theory. Minimal network theory proposes that people make inferences based on the causal model retaining as much as possible of the original causal structure. Thus, in contrast to pruning theory, counterfactuals are not modeled via the do operator. Instead, counterfactual inference involves maintaining the same causal structure and updating the variables in the model so that they are consistent with the counterfactual supposition. In this account, backtracking

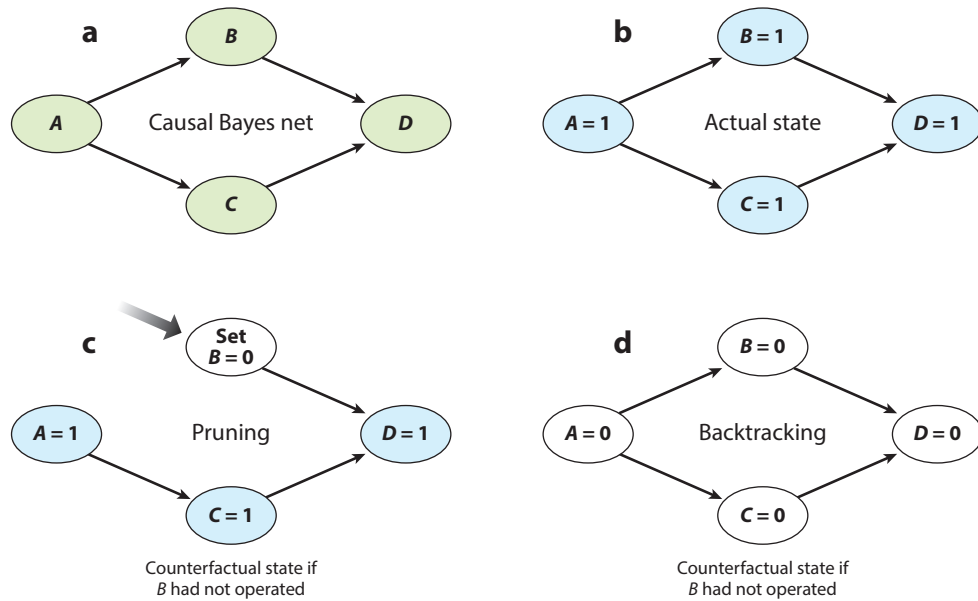


Figure 2

(a) A causal Bayes net representing the causal relations between four components in a mechanical device (each variable can take a value of 1 or 0, corresponding to whether or not it is operating). (b) In the actual world the device is on and all of the components are working (all variables = 1). (c) The counterfactual “If *B* had not operated, would *D* still have operated?” as represented by pruning theory. The counterfactual involves an intervention to set $B = 0$ and to remove the causal link from *A* to *B*. The resultant updating shows that *D* would still be operating. In particular, there is no backtracking from the intervention on *B* to an inference that $A = 0$, so *A* activates *C* and in turn *D*. (d) The same counterfactual as represented by minimal network theory. The counterfactual supposition $B = 0$ does not remove the causal link from *A* to *B*, and a backtracking inference is made that updates *A* to be not operative ($A = 0$). Consequently, all variables, including *D*, change value to 0.

inferences are permitted (see **Figure 2** for an illustration of the differences between the two theories).

Rips & Edwards (2013) test the theories by asking for counterfactual conditional inferences using both deterministic and probabilistic arguments. Results provide more support for minimal network theory than for pruning theory though are not entirely consistent with either, prompting the investigators to propose a reduced version of Hiddleston’s theory. The overall pattern of results suggests that people sometimes do and sometimes do not represent counterfactuals using the do operator.

Lucas & Kemp (2012) propose an extension of Pearl’s (2000) representation to accommodate people’s inconsistency in backtracking. They propose that people carry a causal network to represent the current state of affairs as well as an augmented twin network. One network is used to reason from intervention (using the equivalent of the do operator) and the other from observation. The model provided good fits to published data on counterfactual backtracking.

Gerstenberg et al. (2013) provide support for the idea that people sometimes make inferences via pruning and sometimes do not. They found that the likelihood of making an inference consistent with pruning in a four-variable causal model (such as that depicted in **Figure 2**) depended on the order in which inferences were made. Specifically, if a question about the cause of the counterfactually intervened-on variable was asked early in a sequence of questions, then people were

less likely to prune than if questions about effects were asked early instead. Drawing participants' attention to the cause made them more likely to assign it a value consistent with the counterfactual assumption. It may be that the effect of pruning is not so much to render causes independent of their effects but to render them irrelevant to the discourse. Because they are normally irrelevant, their values do not change. But if attention is brought to them, they must be given a value consistent with the values assumed in the discourse. Dehghani et al. (2012) relatedly conclude that people sometimes backtrack to a cause because of their desire and capacity to explain the counterfactual state of affairs.

DECISION MAKING AND ATTRIBUTION

Decisions concern actions that lead to consequences. In that sense, decision making is all about causal analysis. This has been neglected because of the dominance of the gambling metaphor in the study of decision making. Games of chance are designed to minimize the ability to determine outcomes through causal analysis. Nevertheless, philosophers have been aware of the importance of causality for a long time, and several causal decision theories have been developed (Joyce 1999, Nozick 1969). The fundamental idea of those theories is that choice is an intervention.

Consider the following example from Hagmayer & Sloman (2009): Recent research has shown that of 100 men who help with chores, 82 are in good health, whereas only 32 of 100 men who do not help with chores are in good health. Imagine that a friend of yours is married and is concerned about his health. He read about the research and asks for your advice on whether he should start to do chores to improve his health. What is your recommendation? Should he start to do the chores or not?

In one condition, the correlation between doing the chores and health was explained as a result of direct causal relation: "Doing the chores is an additional exercise every day." In this case, the best advice would be to start doing the chores because it would improve health. But in a second condition, participants were told, "Men who are concerned about equality issues are also concerned about health issues and therefore both help to do the chores and eat healthier food." In other words, the correlation was due to a common cause, not a direct cause. Now the interventional analysis suggests that doing the chores should not be recommended because doing so would render the action independent of its normal cause (a concern about equality). Doing the chores would not be diagnostic of attitudes and thus would have no bearing on health. In this study, three times as many participants recommended action in the direct cause than in the common cause condition. So people have a sense of the logic of intervention and employ it in a decision-making context.

A strong argument can be made that a rational agent should treat choices as interventions: He or she should take all the causes of a choice into account before choice and not learn anything about him or herself from the choice itself (any learning should take place during deliberation and information gathering prior to choice). A psychological theory of causal decision making that spells this out is described in Sloman & Hagmayer (2006).

The difficult issue for such theories is that choices sometimes seem like an observation, not an intervention. An observer of someone else's choice surely learns something about the person's habits and motivations by watching them choose, and sometimes our choices are driven by unconscious processes that lead us to be observers of our own choices. There is also direct evidence that people do not conform to the logic of intervention. In cases of diagnostic self-deception (Chance et al. 2011, Mijovic-Prelec & Prelec 2010, Quattrone & Tversky 1984, Sloman et al. 2010), people perform some action precisely to provide evidence to themselves that they have a positive attribute. For instance, Quattrone & Tversky's (1984) study participants held their hands

in cold water for a length of time that provided supporting evidence that they had a good type of heart. Self-handicapping has a similar property (e.g., McCrea et al. 2008). These cases exhibit the puzzle of agency, that there is no objective way to determine the cause of one's action, whether it be will, an unconscious process, or some external force (cf. Wegner 2002). So sometimes we construe the operative force in a self-serving way, in a way that supports a narrative that we would like to believe. The evidence suggests that we do not always know we are doing so.

Deciding the Actual Cause

When thinking about causality, we care both about general claims (e.g., Does asbestos exposure cause cancer?) and singular claims (e.g., Did Fred's exposure to asbestos cause his cancer?). The latter question of actual causation is pervasive in law, medicine, sports, and elsewhere and is a prerequisite to judgments of responsibility and blame. People make actual cause judgments with relative ease, even when there are competing causes, and yet it is notoriously hard to provide a formal account of how they do so. Philosophers have offered various proposals (e.g., Lewis 1986, Woodward 2003), sharpening the debate without delivering any consensus. Psychological research provides a range of empirical findings, mainly from the field of social psychology, but with no properly developed formal theory [although Kelley's (1973, 1983) use of covariation and causal schemas anticipates the causal Bayes net approach].

Broadly speaking, attempts to define actual causation fall into two categories. (a) Dependency or difference-making accounts involve singular claims that are cashed out in terms of counterfactuals. The central claim is that judgments of actual causation are determined by a comparison between what actually happened and what (counterfactually) could have happened (especially whether the effect would have occurred if the cause had not). (b) Process or transference accounts involve singular claims that are understood in terms of the transfer of some conserved quantity such as force or information (Dowe 2000, Salmon 1994). On this view, *A* causes *B* only if *A* transfers some quantity to *B*.

In the standard counterfactual definition of causation (Lewis 1986), *A* causes *B* if and only if both *A* and *B* occur, and if *A* had not occurred, *B* would not have occurred. This simple version suffers from several problems, including overdetermination and preemption (Paul & Hall 2013). For example, consider two kids—Suzy and Billy—throwing rocks at a bottle (see **Figure 3a**). Suppose that both kids hit the bottle at the same time and the bottle smashes. Moreover, either throw was sufficient to break the bottle. According to the above definition, neither kid's throw is counted as a cause, because neither satisfies the counterfactual condition. If Suzy hadn't thrown a rock, the bottle would still have broken as a result of Billy's throw. So Suzy is not a cause, and Billy is not either for a parallel reason. The outcome is overdetermined. But clearly Suzy and Billy are causes.

A further problem is preemption. Suppose that Suzy throws her rock first and it smashes the bottle (see **Figure 3b**). Billy's throw was on target, too, and it would have hit the bottle if Suzy's hadn't. Here again the counterfactual condition fails to hold: The bottle would still have shattered even if Suzy hadn't thrown. But Suzy is clearly a cause of the bottle breaking.

The counterfactual condition can be modified to handle such cases (e.g., Halpern & Pearl 2005; see below). However, these examples suggest that something is missing in a pure dependency account: the actual causal process linking cause to effect. In the case of preemption, there is an active process from Suzy's throw to the bottle shattering, but none from Billy's throw. This active process is the fundamental condition for process or transference accounts. Empirical studies of these kinds of examples suggest that people prefer to assign causality when some process can be identified and that the counterfactual condition is neither necessary nor sufficient for

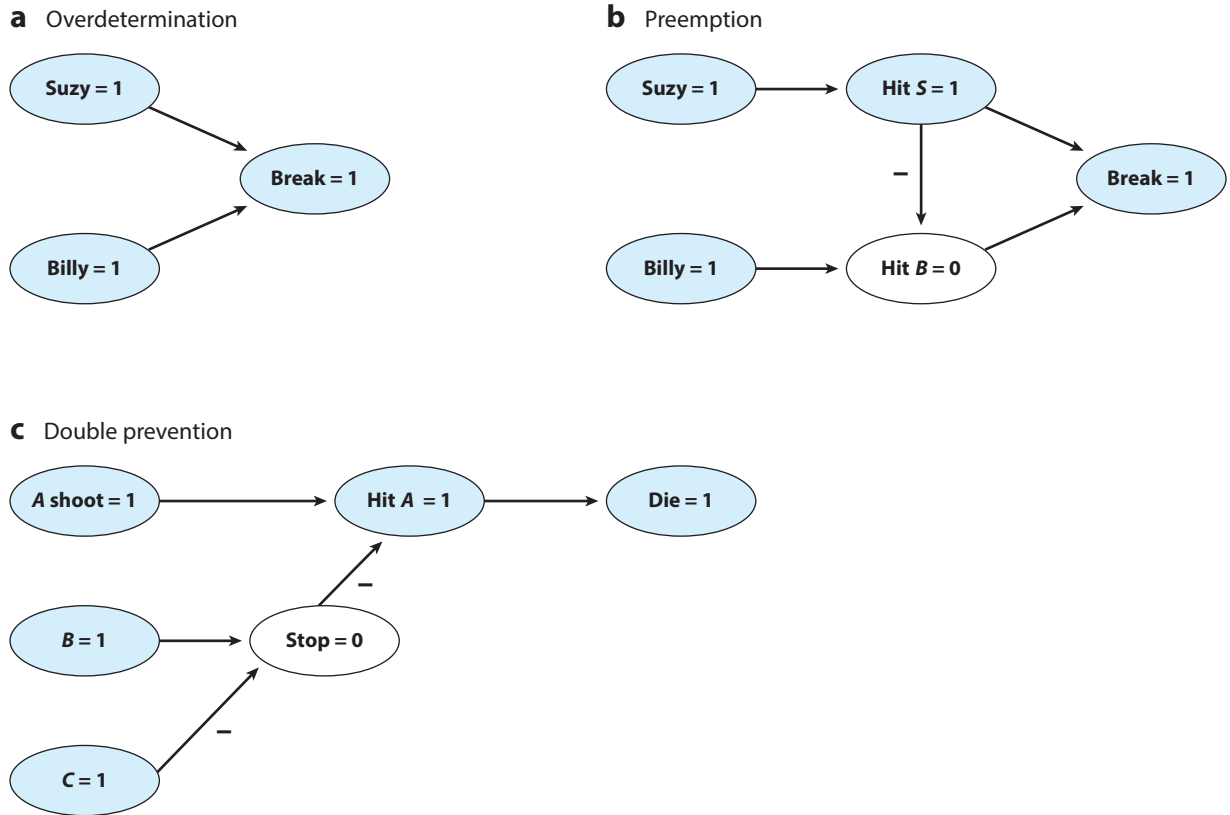


Figure 3

(a) In a representation of overdetermination, both Suzy and Billy throw a rock at the bottle. They hit the bottle at the same time, and it breaks. Either throw on its own would have been sufficient to break the bottle. A simple counterfactual analysis rules that neither is a cause. (b) Preemption: Suzy’s rock hits the bottle first, and the bottle breaks. Billy’s throw misses but would have broken the bottle if Suzy’s throw had not. A simple counterfactual analysis rules that Suzy is not a cause. (c) Double prevention: Assassin *A* shoots at the victim, and bodyguard *B* runs to protect the victim, but his attempt to prevent the shot hitting the victim is thwarted by someone else (*C*), who trips the bodyguard.

assignments of causality. People do not automatically assign causality when the counterfactual condition is met (Walsh & Sloman 2011); moreover, they do assign causality when it is not met, such as in overdetermination and preemption cases (Mandel 2003, Spellman & Kincannon 2001). Furthermore, Walsh & Sloman (2011) show that people’s causal judgments align with transference more than dependence when these are pitted against one another [although Chang (2009) shows that dependent events are judged causal in the absence of transference]. More evidence that we are sensitive to the process by which cause leads to effect comes from causal language. Many causal verbs describe how causes occur and how they have their effects (see Wolff 2013). Wolff & Song (2003) champion a force-dynamics view that is capable of capturing the semantics of many verbs.

Defining causation in terms of process has its own problems (Paul & Hall 2013): It denies causality unless a physical process links putative cause to effect. But many of our causal claims involve omissions or failures to act (or, in the case of physical events, something not occurring). Consider

a prankster who pulls away Granny's chair just before she sits down. Granny falls painfully. What caused the fall? Intuitively it was the prankster's action, even though no active process linked this action to the fall. This example and numerous others (see Shaffer 2012) suggest that our intuitive judgments cannot be explained purely in terms of process theories either.

Lombrozo (2010) provides a partial rectification. Using preemption and more elaborate double prevention cases (see **Figure 3c**), she shows that both types of relation can influence people's judgments depending on whether events are construed teleologically or mechanistically. In particular, with teleological events (e.g., intentional actions) people focus more on dependency, but with physical events they focus more on process. What differentiates teleological contexts is that an agent is aiming for a goal. The exact manner in which this is achieved is less important than whether it is achieved.

Heider (1958) distinguished intentional (or personal) causation from physical causation. For him, an agent's causation is augmented if the agent acts intentionally and with foresight of the consequences (Shaver 1985). In support of these claims, Lagnado & Channon (2008; also Hilton et al. 2010) showed that people rated intentional actions as more causal than equivalent unintentional or physical events. For example, a wife who intentionally gives her husband the wrong pills is rated as more a cause of his death than if her action was accidental. Actions conducted with foresight were also rated as more causal. Our intuitive causal ascriptions in social contexts incorporate both motivational and epistemic factors.

Legal Decision Making

As well as judging what causes what in simple cases, people routinely confront more demanding situations in which they need to identify causes from a large mass of information. A paradigm example of this is in legal reasoning, where the fact finder (often a juror) needs to judge whether the defendant is guilty of a crime on the basis of a large body of evidence. The dominant psychological model, the story model (Pennington & Hastie 1986, 1992), claims that people make decisions by constructing narratives. These narratives rely on rich causal knowledge about both physical and mental causation (for converging findings, see Klein 1999). People construct stories to fit their assumptions about the goal-directed actions of intentional agents, and this is aided by sequencing events into a meaningful causal and temporal process.

The story model thus draws on a mixture of narrative and dependency-based knowledge. The latter provides the generic causal knowledge that allows people to craft specific stories—threads of actual causes that lead to a verdict. With its focus on intentional causality and narrative, the story model captures an important dimension of how people think about causes. However, it leaves out a crucial component of legal reasoning. As well as constructing plausible stories, jurors need to evaluate how well the evidence supports the various stories and assess the strength, credibility, and reliability of the evidence (Schum 1994). Kuhn et al. (1994) suggest that legal reasoners fall on a continuum, from those who satisfice by picking a single story to those who engage in theory-evidence coordination, assessing how well evidence supports different hypotheses and stories as well as the reliability of the evidence (including the credibility of witnesses). Kuhn and colleagues do not provide a normative framework. Causal Bayes nets are a natural candidate (see Fenton et al. 2013, Taroni et al. 2006), capturing the dependency and relevance relations characteristic of legal proof. Moreover, there is evidence that people's reasoning fits with qualitative prescripts of such networks (Lagnado 2011b, Lagnado et al. 2012). This is not to undermine the central role played by story structures but merely to show that they are not sufficient.

Attributing Responsibility

A key role for causal thinking is in the attribution of responsibility—either to oneself or to other people. When blaming or crediting people for their actions we trace a causal path from actions (or inactions) to outcomes of interest. Often these attributions will be straightforward (if Suzy alone throws a rock, breaking the bottle, then she clearly caused the bottle to break and should be held responsible). But sometimes situations are less clear-cut. Perhaps Suzy and Billy both threw rocks at the same time, so their relative contributions are not obvious. More complicated still are cases where several agents combine in a complex task, such as a team of managers and employees completing a job. How is responsibility attributed then? Our knowledge of the causal structure of the situation, and other factors such as the intentions, skill, and knowledge of the players, seems crucial to how we distribute praise or blame (cf. Lagnado et al. 2013, Malle & Knobe 1997, Weiner 1995).

Heider (1958) and Kelley (1973, 1983) emphasized the crucial role of causal knowledge in responsibility attribution. Since then, others have focused on covariation, counterfactuals, and abnormality (Hilton & Slugoski 1986, Mandel 2003, Spellman 1997). But the research lacks a suitable framework to model complex attributions. Here again causal Bayes nets present a possible launch pad. Halpern & Pearl (2005) present a causal Bayes net theory of singular causation closely tied to counterfactual analysis. The key idea is that people use general causal knowledge to assemble a causal model and then evaluate actual causal claims based on counterfactuals derived from this model. To avoid problems of overdetermination and preemption, Halpern & Pearl (2005) propose: (a) to relax the counterfactual condition so that overdetermined causes are still counted as causal and (b) to expand the causal model, where necessary, to include additional variables that capture an active causal process (for supporting data, see Ahn & Bailenson 1996).

For example, when Suzy and Billy both hit the bottle simultaneously, and each throw is sufficient to smash the bottle, proposal *a* allows us to consider the counterfactual dependency between each throw and the bottle smashing contingent on one of the kids not throwing. So Suzy is ruled in as a cause, because if one sets Billy to not throwing, the bottle's smashing depends on Suzy's throw. A similar argument applies to Billy's throw, which is also ruled in as a cause. More generally, Halpern & Pearl (2005) define actual causation such that *A* causes *B* if and only if *B* counterfactually depends on *A* under some suitably constrained set of interventions. To deal with cases of preemption, Halpern & Pearl elaborate the causal model of the situation to capture the notion of a causal process (see **Figure 3b**). For them, this amounts to a more fine-grained dependency analysis.

One benefit of this approach is that it yields a graded measure of causal responsibility (Chockler & Halpern 2004). In short, an agent's responsibility is inversely proportional to the number of changes (in terms of changing values of variables) required to make the agent's action pivotal for the outcome. Thus, Suzy receives greater causal responsibility when she alone hits the bottle than when both she and Billy hit it.

A graded measure of responsibility proves invaluable in group contexts where individuals might differ in how much they contribute to the final outcome. This model of causal responsibility was tested in a series of experiments (Gerstenberg & Lagnado 2010, Zultan et al. 2012). People attributed responsibility in team games, where the contributions of individual team members were combined according to different causal functions to yield the group outcome. In one scenario participants were judges in a cooking competition. Each chef prepared an individual dish, and the overall team result was determined by a function of the individual contributions (e.g., with a conjunctive function, all chefs needed to pass with their individual dish for the team to win; in a disjunctive function, only one chef needed to pass). In such tasks overdetermination is commonplace. Across a wide variety of structures and tasks people's attributions were well fit by the

causal responsibility model derived from Chockler & Halpern (2004). In follow-up studies, however, Lagnado et al. (2013) showed that people's attributions also depended on how much a team member was deemed critical for the team winning before the task was undertaken. Overall, this line of research illustrates the fruitfulness of using causal Bayes nets to explore causal aspects of attribution, but it also shows that there are complexities in human attributions not yet captured by these models.

Moral Reasoning

Causal attribution is necessary in all assignments of moral responsibility, blame, and punishment (Cushman & Young 2011). It would be very odd—or at least wrong—to blame or punish someone for a bad outcome that was causally unrelated to that person's actions. This is so obvious that we rarely think about it explicitly. Causal inference can occur quickly and effortlessly. Observations like this led Sloman et al. (2009b) to conclude that causal structure serves as the infrastructure for moral judgment, a conclusion that converges nicely with Cushman's (2013) theory of moral judgment and blame. In fact, the relation between morality and causality is so close that Knobe (2009) found that people made different causal attributions for an action depending on whether the action was good or bad.

Waldmann & Dieterich (2007) go so far as to argue that some of the fundamental phenomena of moral judgment depend on causal structure. They argue that people believe that agents of harm are more morally culpable than those who cause harm as a side effect and provide intriguing evidence that this effect is mediated by who is the focus of attention in a causal structure.

Wiegmann & Waldmann (2014) propose and validate a theory of transfer between moral intuitions that depends on causal structure. According to the theory, causal models underlying the target and analog moral scenarios mediate transfer effects between them. Mapping occurs between corresponding roles in the two causal structures. Focusing on inductive inference, Lee & Holyoak (2008) also showed that people use causal models to assess the strength of analogical inferences. They report that the coherence of the causal relation was more important for induction than similarity.

INTERIM CONCLUSION

In sum, work on decision making and attribution, like work on reasoning, showcases the importance of causal inference. But it suggests that causal inference is not merely a way of representing and updating probabilities; it is not merely Bayesian inference. Human causal inference involves the construction of narratives that unfold over time and determine the focus of attention, narratives that reflect knowledge of the specific mechanisms that drive effects.

ASYMMETRY OF CAUSE AND EFFECT

If causal relations were grounded in probability relations, there would be no essential difference between reasoning from cause to effect (causal inference) and reasoning from effect to cause (diagnostic inference). For one, they should be equally difficult. The two might differ on a particular occasion because the variables contain different amounts of information about one another; i.e., $P(\text{effect} \mid \text{cause})$ might be different than $P(\text{cause} \mid \text{effect})$, but reasoning should not systematically favor one over the other. Yet it does. Fenker et al. (2005) found that people were faster to verify a causal relation when presented in the cause-to-effect order than when presented in the effect-to-cause order. The asymmetry did not hold when participants were asked to make judgments of

association, which suggests that the greater ease of the causal direction was not due to other factors such as frequency or familiarity. Greater facility with causal than diagnostic reasoning has been found elsewhere. Tversky & Kahneman (1974) showed that people judge $P(\text{effect} \mid \text{cause})$ higher than corresponding $P(\text{cause} \mid \text{effect})$ even in cases in which there would be no reason to expect a difference. Hong et al. (2005) found that children performed better on cause-effect inferences than on effect-cause inferences. There are also strong asymmetries favoring the causal direction in the perception of causality (reviewed by White 2006).

These asymmetries seem to reflect a system that is biased to reason from cause to effect rather than effect to cause. Tversky & Kahneman (1974) suggested that people are able to run mental simulations from cause to effect, mental simulations that travel forward in time. Simulations from effect to cause would be inconsistent with the process being simulated; processes necessarily move forward in time. In this sense, diagnostic reasoning requires a mental facility beyond that of forward causal reasoning. It requires more than the ability to represent relations in the world; it requires the ability to reason about them in a way that is divorced from how they unfold.

The difficulty of diagnostic reasoning is a contentious issue in comparative cognition. Penn et al. (2008) argue that nonhuman animals are incapable of representing the kind of abstract relations that support analogical inference. They argue that diagnostic reasoning depends on such abstract relations because it requires explanatory reasoning about naïve theories and hidden causes, the kinds of reasoning that people are capable of (e.g., Keil 1989, Murphy & Medin 1985, Waldmann & Holyoak 1992). Penn et al. (2008) argue that nonhuman animals are not capable of this sort of diagnostic reasoning. They may be very sophisticated in their ability to respond effectively to circumstance, but all such responses are based on sophisticated associative mechanisms, not causal reasoning.

There is some evidence that challenges this view. Three out of six New Caledonian crows tested were able to reason causally and diagnostically in order to get food out of a special tube outfitted with holes (Taylor et al. 2009). The birds even generalized their solution to a very different apparatus. There's little evidence that primates can accomplish these tasks; too many alternative associative explanations exist for their behavior. But even crows do not appear to be capable of the richest kind of diagnostic reasoning, in which an entirely abstract mechanism is induced and generalized to a completely different problem. Of course, humans do not always shine either, but at least some humans, some of the time, engage in sophisticated diagnostic reasoning. The most we can say about nonhuman animals is that there is weak evidence that crows (and no other animals) can engage in weak diagnostic reasoning that involves associating actions with fairly abstract perceptual cues.

In humans, one type of evidence indicating that causal inference involves different mechanisms than diagnostic inference is that the two are differentially sensitive to alternative causes. Consider the following question: "The mother of a newborn baby has dark skin. How likely is it that her baby has dark skin?" In answering the question, one should consider the possibility that the baby inherited dark skin from the mother but also the possibility that the baby inherited dark skin from the father or via any other route. Alternative causes can only raise the probability of an effect, so the answer to this question should be higher than the answer to the question, "How likely is it that a baby has dark skin inherited from its mother?" But it's not. Fernbach et al. (2010, 2011a) found that people neglect alternative causes when judging the probability of an effect given its cause. This may be because people imagine the transmission from cause to effect, and such a mental simulation does not offer a way to consider alternative causes.

People do not, however, neglect alternative causes when making diagnostic judgments. Consider the probability that a mother has dark skin given that her baby does. This should be lower than the probability that the mother has dark skin given that the baby does and the baby does not

have dark skin for any other reason. Indeed, the latter probably should be 1 (or close to it), and people do in fact judge it high—higher than the first probability, in which the state of alternative causes is unknown. Thus in this case diagnostic reasoning is more accurate than causal reasoning.

The fact that people neglect alternative causes in causal reasoning implies that people will ignore more likely causes when presented with less likely ones. In particular, providing weak evidence in the form of a weak cause should lead people to think the effect is less likely than they would have if the cause had not been presented, as long as they would have thought of a stronger cause themselves. Fernbach et al. (2011b) demonstrated this in two different paradigms. In one, people were asked if they wanted to bet on the outcome of a congressional election. They could either choose \$10 for sure or \$30 if the Republicans won the house in an upcoming election (they would get their selection with a fixed probability). In one case, they were told that a newspaper had endorsed a local Republican candidate; in the other, they were told nothing. Everyone agreed that the endorsement made a win by the Republicans more likely, but not by much. It was very weak evidence. The results were that twice as many people bet on the election given no evidence other than the weak positive evidence. Apparently, when presented a target cause, people neglect alternative causes even when the target cause is weak, so weak that it would be easy to generate a stronger cause oneself.

All of these studies on the asymmetry of causal reasoning are consistent with the idea that people are much better able to run mental simulations forward from cause to effect than backward from effect to cause. Mental simulations depend on more information than mere probabilistic dependency. They require instantiating the critical variables and in doing so elicit concrete information; they cannot be purely abstract (Barsalou 1999). They take place over time, and the speed of the simulation can represent the speed with which the events themselves occur. Hegarty (2004) has studied how people think about systems of cogs, enmeshed gears turning. Schwartz & Black (1999) find that cogs don't all move together in imagination; rather, people think about them one after the other. Finally, simulations tend to involve a representation of *how* the effect occurs, not just *that* it occurs. In that sense, simulations require more than an understanding of dependency. They also require an understanding of the mechanism by which causes bring about effects.

Causal inference suffers from other sorts of neglect. When using 2×2 contingency tables to garner evidence for a causal relation, people tend to focus on the number of cases in which both cause and effect are present and neglect other cells (Shaklee & Goldston 1989; reviewed by Hattori & Oaksford 2007; cf. Laux Goedert & Markman 2010). Kahan et al. (2013) found that when cause and effect were neutral—a skin rash treatment, variables that the reasoner had no prior beliefs about—people highest in numeracy did better than peers who were less numerate. However, when the same data were presented as the results of a study of a gun-control ban, the more numerate participants were more biased in their conclusions and indeed more polarized in attitude than were the less numerate. The skill to analyze and interpret data proved helpful when the information was neutral, but it proved helpful in a different sense when the information mattered to the person: It provided the means to come to a biased conclusion.

MARKOV VIOLATIONS

The Markov property of Bayes nets is a relation between a graph and a probability distribution. If the Markov property holds, the graph faithfully represents the relations of dependence and independence implied by the probability distribution. For instance, a causal chain from A to B to C implies that A , B , and C are all dependent on one another but that A and C should be independent if the value of B is known. A number of experiments have examined whether people conform to the Markov property by describing a set of causal relations and then asking for probability judgments.

In the *ABC* chain for instance, people should judge $P(C \text{ is high} \mid A \text{ is high}, B \text{ is high})$ to be the same as $P(C \text{ is high} \mid A \text{ is low}, B \text{ is high})$. In most such experiments, many violations of the Markov property have been observed (Chaigneau et al. 2004, Lagnado & Sloman 2004b, Mayrhofer et al. 2008, 2010, Rehder & Burnett 2005, Waldmann & Hagmayer 2005, Walsh & Sloman 2008).

Park & Sloman (2013) varied whether the mechanisms underlying the two causal relations in a causal chain and common cause structure were the same or different. Violations only occurred when the mechanisms were the same. They explain this by proposing that when people observe an unexpected outcome such as a cause occurring without its effect (Park & Sloman 2014), they explain the discrepancy by positing hidden variables in the causal structure such as disablers that had not been mentioned. When mechanisms are the same, a single disabler can disable all outcomes related to a single cause and provide an inferential route between the two effects. When mechanisms are different, the disabler for one effect should not influence another effect and thus no new route is available. We conclude that people do represent causal relations in a way that is consistent with the Markov property; however, people do not necessarily represent only the relations that the experimenter mentions. They construct a causal graph that is consistent with the mechanism they imagine to hold. Mayrhofer & Waldmann (2014) argue further that people take the variable's disposition into account. They observe that Markov violations are more common when, for example, the cause—as opposed to the effect—is an agent in a common cause model.

The fact that people elaborate representations based on mechanistic knowledge aligns well with our conclusion from work on cause-effect asymmetry that people (and nonhuman animals) make causal inferences using mental simulation. Information about how mechanisms operate is precisely what one needs to mentally simulate the unfolding of a process over time.

ILLUSION OF EXPLANATORY DEPTH

We do not always understand mechanisms in great detail. Indeed, people tend to think we understand causal systems better than we actually do. Rozenblit & Keil (2002) demonstrated this by asking people how well they understood everyday objects such as speedometers and ballpoint pens. They then asked people how the objects actually worked; they asked for a causal explanation of the device. In most cases participants failed to come up with an adequate explanation, and when asked again to rate their level of knowledge, the average rating was lower. In other words, the inability to generate a causal mechanism led people to the conclusion that their own previous sense of understanding was inflated; they were suffering from an illusion of understanding. Rozenblit & Keil (2002) also showed that the effect of explanation does not occur for facts (e.g., capitals of countries), procedures (e.g., how to make an international phone call), or narratives (e.g., plots of familiar movies), which suggests that the effect is not a special case of overconfidence. The effect applies to people's sense of understanding causal systems and also to argument justification (Fisher & Keil 2013). Fernbach et al. (2013) demonstrate that the illusion arises for political policies and that puncturing the illusion through explanation not only reduces people's sense of understanding, it also moderates the extremity of their political positions.

The illusion of understanding may be due to mental distance from the objects of judgment. The effect is reduced when people are induced to assume a lower level of construal (Alter et al. 2010). Relatedly, Keil (2006) finds that the effect is greatest for systems whose components are visible. The concreteness of the image associated with such objects fosters a strong sense of understanding. This is consistent with Lawson's (2006) findings that the illusion of understanding is very strong for bicycles even among bicycle experts. She found that errors persisted, though in weaker form, in the presence of a real bicycle.

These findings provide some support for the causal Bayes net assumption that people do not represent complete mechanisms. Rather, most knowledge is sparse and abstract. The fact that knowledge can be fleshed out implies that mental representation has a hierarchical character such that abstract knowledge can be unpacked into detailed knowledge under the right conditions (Goodman et al. 2011, Griffiths & Tenenbaum 2009). Hierarchical Bayes nets provide a formal model for representing knowledge that has multiple levels of abstraction (Kemp & Tenenbaum 2009, Tenenbaum et al. 2011). Such models do not clearly capture another property of causal knowledge: that it often includes black boxes (Strevens 2008) that point to details that are missing without spelling them out. Somehow our causal representations refer to details that are not themselves represented by the individual but are available in the community, often by appealing to experts (Keil 2012).

DISCUSSION

We share with Holyoak & Cheng (2011) the view that causal Bayes nets have emerged as the leading normative theory of causal learning and reasoning. The virtues of a causal Bayes net are its generality and the utility and apparent veridicality of its representations. The framework has provided solutions to several long-unsolved problems. For example, it has been an open question how people choose the possible world (or worlds) to reason about when confronted with a conditional antecedent or counterfactual assumption. Clearly the assumption should be taken as true (Ramsey 1929/1990), but what else should? As Lewis (1973) puts it, Which possible world is the most similar to the actual world? We believe that Pearl (2000) provides the best answer: It is the world in which the normal upstream causes of the antecedent or counterfactual assumption retain their actual values, but causal inference is used to update the values of the downstream effects of the assumed variables. This model is not universally applicable, but it does appear to account for many cases (for a review, see Sloman & Pearl 2013).

But the causal Bayes nets framework has not offered a silver bullet that answers all questions about human thought. We have seen that it is not always clear whether a choice should be modeled as an intervention. There are also problems with deriving a normative model of actual causation. The framework also suffers from its generality. It is powerful enough that some model can usually be found to fit data after the fact.

On Mechanism

The major shortcoming of the framework in our view is its incompleteness: It fails to represent all the properties of mechanisms, specifically the process and timing by which a causal event unfolds (Cartwright 2004). In some cases, this means that the framework appears to make incorrect predictions, as our discussion of the Markov property reveals. Researchers reported many violations of this property until it was recognized that the violations did not occur with respect to a model that takes into account information about mechanism that goes beyond probabilistic dependency. The study of causal attribution similarly revealed that people consider an active causal process, not just what depends on what. There are plenty of other findings supporting our appeal to mechanism, from data on language (Wolff 2013) to adult and developmental learning data (Rottman & Keil 2012, Rottman et al. 2014, Schlottmann 1999). People are surprisingly weak at inferring causal structure from correlational data and instead use intervention, temporal delays, and instruction (Buehner & May 2002; Lagnado & Sloman 2004b, 2006). People are especially sensitive to spatiotemporal information.

Our analysis of decision making finds that people depend on what may be a different kind of information that goes beyond dependency, namely narrative structure. And our analysis of causal

asymmetries suggests that people have a special tool—mental simulation—for reasoning from cause to effect (as opposed to reasoning from effect to cause).

We suspect that these three findings—the importance of mechanism, the importance of narrative, and the use of mental simulation—are closely related. Mental simulations are representations over time of mechanisms. When multiple actors are involved, those simulations become more complex, like little dramas, and we refer to them as narratives. The ability to construct these complex mental simulations may be the source of the human tendency to tell stories. Stories are shared representations of how causes lead to effects when multiple actors are involved.

So the causal Bayes nets framework is a productive one, perhaps even more productive than the probability framework (see Chater & Oaksford 2013), although incomplete. It has many implications that have not fully played out. For instance, theorists have been aware for a long time of the importance of causation to categorization (Keil 1989, Lewis 1929, Murphy & Medin 1985) and to induction (Lassaline 1996). Rehder (2003) integrated these ideas into a formal causal model that combines prior causal knowledge with observed properties to form categories and make probabilistic inferences. This approach successfully models how people classify novel instances of artificial categories and make inductive inferences about them. Along with Shafto et al. (2008), Rehder (2006) shows that causality trumps similarity in induction. Rehder's model has been extended in multiple ways. For instance, Hayes & Rehder (2012) show that children depend more on causal coherence in their classification behavior than adults do.

Causal knowledge is a tool for categorization in that it can help to fill in missing information (inferring unobserved features, making predictions, or determining the functionality or coherence of an object). Of course, causal structure can only relate features; features themselves are necessary to determine the identity of an object. Puebla & Chaigneau (2014) find that people categorize artifacts by features if the features are known and use causal structure to make inferences if all features are not known.

What is missing from the causal Bayes nets framework in its current state is a specification of the role of mechanism. When knowledge is abstract, mechanistic details are presumably not important. But when modeling expertise, it will surely be necessary to develop models that do more than represent dependencies; models must also represent how processes unfold over time.

CONCLUSION

We believe the causal Bayes nets framework is incomplete in that it does not offer a principled basis for modeling how people think about mechanisms. As a descriptive model, it is incorrect in other ways as well. Bayes nets are representations of probability, and we have known for over 40 years that people make systematic errors in their judgments of probability (Edwards 1968, Kahneman et al. 1982). So it is not surprising that people make systematic errors when using causal structure to judge probability (Bes et al. 2012, Fernbach et al. 2011a). We are impressed by people's sensitivity to qualitative causal structure (e.g., Cummins 1995) but less impressed by their quantitative acumen.

The fundamental problem, we believe, is that causal Bayes nets lend themselves to two interpretations: a view from the outside, in which they are representations of probability, and a view from the inside, in which they are (partial) representations of mechanistic interactions. Psychologically, if not normatively, these views are different, and the formalism does not help make the psychological distinction. What people care about when they think causally, beyond just patterns of dependency, is a narrative-like unfolding of events over time. The process of mental simulation helps to construct such narratives. It supports much of our reasoning but currently eludes a strict formal analysis.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

- Ahn W, Bailenson J. 1996. Mechanism-based explanations of causal attribution: an explanation of conjunction and discounting effect. *Cogn. Psychol.* 31:82–123
- Ali N, Chater N, Oaksford M. 2011. The mental representation of causal conditional reasoning: mental models or causal models. *Cognition* 119(3):403–18
- Alter AL, Oppenheimer DM, Zemla JC. 2010. Missing the trees for the forest: a construal level account of the illusion of explanatory depth. *J. Personal. Soc. Psychol.* 99(3):436–51
- Barbey AK, Sloman SA. 2007. Base-rate respect: from ecological rationality to dual processes. *Behav. Brain Sci.* 30:241–54
- Barsalou LW. 1999. Perceptions of perceptual symbols. *Behav. Brain Sci.* 22(4):637–60
- Bechlivanidis C, Lagnado DA. 2013. Does the “why” tell us the “when”? *Psychol. Sci.* 24:1563–72
- Bes B, Sloman SA, Lucas CG, Raufaste E. 2012. Non-Bayesian inference: Causal structure trumps correlation. *Cogn. Sci.* 36:1178–203
- Buehner MJ, Humphreys GR. 2009. Causal binding of actions to their effects. *Psychol. Sci.* 20(10):1221–28
- Buehner MJ, Humphreys GR. 2010. Causal contraction spatial binding in the perception of collision events. *Psychol. Sci.* 21:44–48
- Buehner MJ, May J. 2002. Knowledge mediates the timeframe of covariation assessment in human causal induction. *Think. Reason.* 8(4):269–95
- Cartwright N. 2004. Causation: one word, many things. *Philos. Sci.* 71:805–81
- Chaigneau SE, Barsalou LW, Sloman S. 2004. Assessing the causal structure of function. *J. Exp. Psychol.: Gen.* 133:601–25
- Chance Z, Norton MI, Gino F, Ariely D. 2011. Temporal view of the costs and benefits of self-deception. *Proc. Natl. Acad. Sci. USA* 108(Suppl. 3):15655–59
- Chandon E, Janiszewski C. 2009. The influence of causal conditional reasoning on the acceptance of product claims. *J. Consum. Res.* 35(6):1003–11
- Chang W. 2009. Connecting counterfactual and physical causation. In *Proc. 31th Annu. Conf. Cogn. Sci. Soc.*, pp. 1983–87. Austin, TX: Cogn. Sci. Soc.
- Chater N, Oaksford M. 2013. Programs as causal models: speculations on mental programs and mental representation. *Cogn. Sci.* 37(6):1171–91
- Cheng P. 1997. From covariation to causation: a causal power theory. *Psychol. Rev.* 104(2):367–405
- Chockler H, Halpern JY. 2004. Responsibility and blame: a structural-model approach. *J. Artif. Intell. Res.* 22(1):93–115
- Cummins DD. 1995. Naive theories and causal deduction. *Mem. Cogn.* 23(5):646–58
- Cummins DD, Lubart T, Alksnis O, Rist R. 1991. Conditional reasoning and causation. *Mem. Cogn.* 19(3):274–82
- Cushman FA. 2013. Action, outcome and value: a dual-system framework for morality. *Personal. Soc. Psychol. Rev.* 17(3):273–92
- Cushman FA, Young L. 2011. Patterns of moral judgment derive from nonmoral psychological representations. *Cogn. Sci.* 35:1052–75
- Dehghani M, Iliev R, Kaufmann S. 2012. Causal explanation and fact mutability in counterfactual reasoning. *Mind Lang.* 27(1):55–85
- Dowe P. 2000. *Physical Causation*. New York: Cambridge Univ. Press
- Edwards W. 1968. Conservatism in human information processing. In *Formal Representation of Human Judgment*, ed. B Kliemuntz, pp. 17–52. New York: Wiley
- Fenker D, Waldmann MR, Holyoak KJ. 2005. Accessing causal relations in semantic memory. *Mem. Cogn.* 33:1036–46

- Fenton N, Neil M, Lagnado DA. 2013. A general structure for legal arguments about evidence using Bayesian networks. *Cogn. Sci.* 37:61–102
- Fernbach PM, Darlow A, Sloman SA. 2010. Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychol. Sci.* 21(3):329–36
- Fernbach PM, Darlow A, Sloman SA. 2011a. Asymmetries in causal and diagnostic reasoning. *J. Exp. Psychol.: Gen.* 140(2):168–85
- Fernbach PM, Darlow A, Sloman SA. 2011b. When good evidence goes bad: the weak evidence effect in judgment and decision-making. *Cognition* 119:459–67
- Fernbach PM, Erb CD. 2013. A quantitative causal model theory of conditional reasoning. *J. Exp. Psychol.: Learn. Mem. Cogn.* 39(5):1327–43
- Fernbach PM, Rogers T, Fox C, Sloman SA. 2013. Political extremism is supported by an illusion of understanding. *Psychol. Sci.* 24:939–46
- Fisher M, Keil FC. 2013. The illusion of argument justification. *J. Exp. Psychol.: Gen.* 143:425–33
- Gerstenberg T, Bechlvianidis C, Lagnado DA. 2013. Back on track: backtracking in counterfactual reasoning. In *Proc. 35th Annu. Conf. Cogn. Sci. Soc.*, ed. M Knauff, M Pauen, N Sebanz, I Wachsmuth, pp. 2386–91. Austin, TX: Cogn. Sci. Soc.
- Gerstenberg T, Lagnado DA. 2010. Spreading the blame: the allocation of responsibility amongst multiple agents. *Cognition* 115(1):166–71
- Goldvarg E, Johnson-Laird PN. 2001. Naive causality: a mental model theory of causal meaning and reasoning. *Cogn. Sci.* 25:565–610
- Goodman ND, Ullman TD, Tenenbaum JB. 2011. Learning a theory of causality. *Psychol. Rev.* 118(1):110–19
- Griffiths TL, Tenenbaum JB. 2009. Theory-based causal induction. *Psychol. Rev.* 116(4):661–716
- Hagmayer Y, Sloman SA. 2009. Decision makers conceive of themselves as interveners, not observers. *J. Exp. Psychol.: Gen.* 138:22–38
- Halpern JY, Pearl J. 2005. Causes and explanations: a structural-model approach. Part I: causes. *Br. J. Philos. Sci.* 56(4):843–87
- Hampton JA. 2012. Thinking intuitively the rich (and at times illogical) world of concepts. *Curr. Dir. Psychol. Sci.* 21(6):398–402
- Hattori M, Oaksford M. 2007. Adaptive non-interventional heuristics for covariation detection in causal induction: model comparison and rational analysis. *Cogn. Sci.* 31(5):765–814
- Hayes BK, Rehder B. 2012. The development of causal categorization. *Cogn. Sci.* 36(6):1102–28
- Hegarty M. 2004. Mechanical reasoning as mental simulation. *Trends Cogn. Sci.* 8:280–85
- Heider F. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley
- Hiddleston E. 2005. A causal theory of counterfactuals. *Noûs* 39(4):632–57
- Hilton DJ, McClure J, Sutton RM. 2010. Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes? *Eur. J. Soc. Psychol.* 40(3):383–400
- Hilton DJ, Slugoski BR. 1986. Knowledge-based causal attribution: the abnormal conditions focus model. *Psychol. Rev.* 93(1):75–88
- Holyoak KJ, Cheng PW. 2011. Causal learning and inference as a rational process: the new synthesis. *Annu. Rev. Psychol.* 62:135–63
- Hong L, Chijun Z, Xuemei G, Shan G, Chongde L. 2005. The influence of complexity and reasoning direction on children's causal reasoning. *Cogn. Dev.* 20(1):87–101
- Joyce J. 1999. *The Foundations of Causal Decision Theory*. Cambridge, UK: Cambridge Univ. Press
- Kahan DM, Peters E, Dawson EC, Slovic P. 2013. *Motivated numeracy and enlightened self-government*. Yale Law School, The Cult. Cogn. Proj., Work. Pap. 116
- Kahneman D, Slovic P, Tversky A, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge Univ. Press
- Kahneman D, Tversky A. 1982a. The simulation heuristic. See Kahneman et al. 1982, pp. 201–8
- Kahneman D, Tversky A. 1982b. Variants of uncertainty. *Cognition* 11(2):143–57
- Keil FC. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press
- Keil FC. 2006. Explanation and understanding. *Annu. Rev. Psychol.* 57:227–54
- Keil FC. 2012. Running on empty? How folk science gets by with less. *Curr. Dir. Psychol. Sci.* 21(5):329–34

- Kelley HH. 1973. The processes of causal attribution. *Am. Psychol.* 28(2):107–28
- Kelley HH. 1983. Perceived causal structures. In *Attribution Theory and Research: Conceptual, Developmental and Social Dimensions*, ed. JM Jaspars, FD Fincham, M Hewstone, pp. 343–69. New York: Academic
- Kemp C, Tenenbaum JB. 2009. Structured statistical models of inductive reasoning. *Psychol. Rev.* 116(1):20–58
- Klein G. 1999. *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press
- Knobe J. 2009. Folk judgments of causation. *Stud. Hist. Philos. Sci.* 40(2):238–42
- Kuhn D, Weinstock M, Flaton R. 1994. How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychol. Sci.* 5:289–96
- Lagnado DA. 2011a. Causal thinking. In *Causality in the Sciences*, ed. PM Illari, F Russo, J Williamson, pp. 129–49. New York: Oxford Univ. Press
- Lagnado DA. 2011b. Thinking about evidence. *Proc. Br. Acad.* 171:183–223
- Lagnado DA, Channon S. 2008. Judgments of cause and blame: the effects of intentionality and foreseeability. *Cognition* 108(3):754–70
- Lagnado DA, Fenton N, Neil M. 2012. Legal idioms: a framework for evidential reasoning. *Argument Comput.* 4:46–63
- Lagnado DA, Gerstenberg T, Zultan R. 2013. Causal responsibility and counterfactuals. *Cogn. Sci.* 37:1036–73
- Lagnado DA, Sloman SA. 2004a. Inside and outside probability judgment. In *Blackwell Handbook of Judgment and Decision Making*, ed. DJ Koehler, N Harvey, pp. 157–76. London: Blackwell
- Lagnado DA, Sloman SA. 2004b. The advantage of timely intervention. *J. Exp. Psychol.: Learn. Mem. Cogn.* 30:856–76
- Lagnado DA, Sloman SA. 2006. Time as a guide to cause. *J. Exp. Psychol.: Learn. Mem. Cogn.* 32:451–60
- Lassaline ME. 1996. Structural alignment in induction and similarity. *J. Exp. Psychol.: Learn. Mem. Cogn.* 22:754–70
- Laux JP, Goedert KM, Markman AB. 2010. Causal discounting in the presence of a stronger cue is due to bias. *Psychon. Bull. Rev.* 17(2):213–18
- Lawson R. 2006. The science of cycology: failures to understand how everyday objects work. *Mem. Cogn.* 34(8):1667–75
- Lee HS, Holyoak KJ. 2008. The role of causal models in analogical inference. *J. Exp. Psychol.: Learn. Mem. Cogn.* 34(5):1111–22
- Lewis CI. 1929 (1956). *Mind and the World Order: An Outline of a Theory of Knowledge*. New York: Scribner's Sons. Reprint ed.
- Lewis D. 1973. Causation. *J. Philos.* 70(17):556–67
- Lewis D. 1986. Causal explanation. *Philos. Pap.* 2:214–40
- Lombrozo T. 2010. Causal-explanatory pluralism: how intentions, functions, and mechanisms influence causal ascriptions. *Cogn. Psychol.* 61(4):303–32
- Lucas A, Kemp C. 2012. A unified theory of counterfactual reasoning. In *Proc. 34th Annu. Meet. Cogn. Sci. Soc.*, ed. N Miyake, D Peebles, RP Cooper, pp. 707–12. Austin, TX: Cogn. Sci. Soc.
- Malle BF, Knobe J. 1997. The folk concept of intentionality. *J. Exp. Soc. Psychol.* 33:101–21
- Mandel DR. 2003. Judgment dissociation theory: an analysis of differences in causal, counterfactual and covariational reasoning. *J. Exp. Psychol.: Gen.* 132(3):419–34
- Markovits H, Potvin F. 2001. Suppression of valid inferences and knowledge structures: the curious effect of producing alternative antecedents on reasoning with causal conditionals. *Mem. Cogn.* 29(5):736–44
- Mayrhofer R, Goodman ND, Waldmann MR, Tenenbaum JB. 2008. Structured correlation from the causal background. In *Proc. 30th Annu. Conf. Cogn. Sci. Soc.*, ed. BC Love, K McRae, VM Sloutsky, pp. 303–8. Austin, TX: Cogn. Sci. Soc.
- Mayrhofer R, Hagmayer Y, Waldmann M. 2010. Agents and causes: a Bayesian error attribution model of causal reasoning. In *Proc. 32nd Annu. Conf. Cogn. Sci. Soc.*, ed. S Ohlsson, R Catrambone, pp. 925–30. Austin, TX: Cogn. Sci. Soc.
- Mayrhofer R, Waldmann MR. 2014. Agents and causes: dispositional intuitions as a guide to causal structure. *Cogn. Sci.* In press
- McCrea SM, Hirt ER, Hendrix KL, Milner BJ, Steele NL. 2008. The worker scale: developing a measure to explain gender differences in behavioral self-handicapping. *J. Res. Personal.* 42(4):949–70

- Michotte A. 1963. *The Perception of Causality*. London: Methuen
- Mijovic-Prelec D, Prelec D. 2010. Self-deception as self-signalling: a model and experimental evidence. *Philos. Trans. R. Soc. B* 365:227–40
- Morris MW, Larrick RP. 1995. When one cause casts doubt on another: a normative analysis of discounting in causal attribution. *Psychol. Rev.* 102(2):331–55
- Murphy GL, Medin DL. 1985. The role of theories in conceptual coherence. *Psychol. Rev.* 92(3):289–316
- Novick LR, Cheng PW. 2004. Assessing interactive causal influence. *Psychol. Rev.* 111:455–85
- Nozick R. 1969. Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G Hempel*, ed. N Rescher, pp. 114–46. Dordrecht: Reidel
- Park J, Sloman SA. 2013. Mechanistic beliefs determine adherence to the Markov property in causal reasoning. *Cogn. Psychol.* 67:186–216
- Park J, Sloman SA. 2014. Causal explanation in the face of contradiction. *Mem. Cogn.* In press
- Paul L, Hall N. 2013. *Causation: A User's Guide*. Oxford, UK: Oxford Univ. Press
- Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Palo Alto, CA: Morgan Kaufmann
- Pearl J. 2000. *Causality: Models, Reasoning and Inference*. New York: Cambridge Univ. Press
- Pearl J. 2013. Structural counterfactuals: a brief introduction. *Cogn. Sci.* 37:977–85
- Penn DC, Holyoak KJ, Povinelli DJ. 2008. Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31(2):109–30
- Pennington N, Hastie R. 1986. Evidence evaluation in complex decision making. *J. Personal. Soc. Psychol.* 51:242–58
- Pennington N, Hastie R. 1992. Explaining the evidence: test of the story model for juror decision making. *J. Personal. Soc. Psychol.* 62:189–206
- Puebla G, Chaigneau SE. 2014. Inference and coherence in causal-based artifact categorization. *Cognition* 130(1):50–65
- Quattrone G, Tversky A. 1984. Causal versus diagnostic contingencies: on self-deception and on the voter's illusion. *J. Personal. Soc. Psychol.* 46:237–48
- Ramsey FP. 1929/1990. General propositions and causality. In *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. DH Mellor, pp. 145–63. London: Humanit. Press
- Rehder B. 2003. A causal-model theory of conceptual representation and categorization. *J. Exp. Psychol.: Learn. Mem. Cogn.* 29(6):1141–59
- Rehder B. 2006. When causality and similarity compete in category-based property induction. *Mem. Cogn.* 34:3–16
- Rehder B, Burnett RC. 2005. Feature inference and the causal structure of object categories. *Cogn. Psychol.* 50:264–314
- Rips LJ. 2010. Two causal theories of counterfactual conditionals. *Cogn. Sci.* 34(2):175–221
- Rips LJ, Edwards B. 2013. Inference and explanation in counterfactual reasoning. *Cogn. Sci.* 37:1107–35
- Rottman BM, Keil FC. 2012. Causal structure learning over time: observations and interventions. *Cogn. Psychol.* 64:93–125
- Rottman BM, Kominsky JF, Keil FC. 2014. Children use temporal cues to learn causal directionality. *Cogn. Sci.* 38(3):489–513
- Rozenblit L, Keil F. 2002. The misunderstood limits of folk science: an illusion of explanatory depth. *Cogn. Sci.* 26(5):521–62
- Russell B. 1913. On the notion of cause. *Proc. Aristotelian Soc.* 13:1–26
- Salmon WC. 1994. Causality without counterfactuals. *Philos. Sci.* 61(2):297–312
- Schlottmann A. 1999. Seeing it happen and knowing how it works: how children understand the relation between perceptual causality and knowledge of underlying mechanisms. *Dev. Psychol.* 35:303–17
- Schum DA. 1994. *The Evidential Foundations of Probabilistic Reasoning*. New York: Wiley
- Schwartz DL, Black T. 1999. Inferences through imagined actions: knowing by simulated doing. *J. Exp. Psychol.: Learn. Mem. Cogn.* 25:116–36
- Shaffer J. 2012. Disconnection and responsibility. *Legal Theory* 18:399–435
- Shafto P, Kemp C, Bonawitz EB, Coley JD, Tenenbaum JB. 2008. Inductive reasoning about causally transmitted properties. *Cognition* 109:175–92

- Shaklee H, Goldston D. 1989. Development in causal reasoning: information sampling and judgment rule. *Cogn. Dev.* 4(3):269–81
- Shaver KG. 1985. *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. New York: Springer-Verlag
- Sirota M, Juanchich M, Haggmayer Y. 2014. Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychon. Bull. Rev.* 21:198–204
- Sloman SA. 2005. *Causal Models: How People Think About the World and Its Alternatives*. New York: Oxford Univ. Press
- Sloman SA, Fernbach PM, Ewing S. 2009b. Causal models: the representational infrastructure for moral judgment. In *Psychology of Learning and Motivation*, Vol. 50: *Moral Judgment and Decision Making*, ed. DM Bartels, CW Bauman, LJ Skitka, DL Medin, pp. 1–26. San Diego, CA: Academic
- Sloman SA, Fernbach PM, Haggmayer Y. 2010. Self-deception requires vagueness. *Cognition* 115(2):268–81
- Sloman SA, Haggmayer Y. 2006. The causal psycho-logic of choice. *Trends Cogn. Sci.* 18;10(9):407–12
- Sloman SA, Lagnado DA. 2005. Do we “do”? *Cogn. Sci.* 29(1):5–39
- Sloman SA, Pearl J, eds. 2013. Special issue: 2011 Rumelhart Prize Special Issue honoring Judea Pearl. *Cogn. Sci.* 37(6):969–1191
- Spellman BA. 1997. Crediting causality. *J. Exp. Psychol.: Gen.* 126(4):323–48
- Spellman BA, Kincannon A. 2001. The relation between counterfactual (“but for”) and causal reasoning: experimental findings and implications for jurors’ decisions. *Law Contemp. Probl.* 64(4):241–64
- Spirtes P, Glymour CN, Scheines R. 1993/2000. *Causation, Prediction, and Search*, Vol. 81. Cambridge, MA: MIT Press
- Strevens M. 2008. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard Univ. Press
- Taroni F, Aitken C, Garbolino P, Biedermann A. 2006. *Bayesian Networks and Probabilistic Inference in Forensic Science*. Chichester, UK: Wiley
- Taylor AH, Hunt GR, Medina FS, Gray RD. 2009. Do new Caledonian crows solve physical problems through causal reasoning? *Proc. R. Soc. B* 276(1655):247–54
- Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. 2011. How to grow a mind: statistics, structure, and abstraction. *Science* 331(6022):1279–85
- Thompson VA. 1995. Conditional reasoning: the necessary and sufficient conditions. *Can. J. Exp. Psychol.* 49(1):1–60
- Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185:1124–31
- Waldmann MR, Dieterich JH. 2007. Throwing a bomb on a person versus throwing a person on a bomb: intervention myopia in moral intuitions. *Psychol. Sci.* 18(3):247–53
- Waldmann MR, Haggmayer Y. 2005. Seeing versus doing: two modes of accessing causal knowledge. *J. Exp. Psychol.: Learn. Motiv. Cogn.* 31:216–27
- Waldmann MR, Haggmayer Y. 2013. Causal reasoning. In *Oxford Handbook of Cognitive Psychology*, ed. D Reisberg, pp. 733–52. New York: Oxford Univ. Press
- Waldmann MR, Holyoak KJ. 1992. Predictive and diagnostic learning within causal models: asymmetries in cue competition. *J. Exp. Psychol.: Gen.* 121(2):222–36
- Walsh CR, Sloman SA. 2008. Updating beliefs with causal models: violations of screening off. In *Memory and Mind: A Festschrift for Gordon H. Bower*, ed. MA Gluck, JR Anderson, SM Kosslyn, pp. 345–57. Hillsdale, NJ: Erlbaum
- Walsh CR, Sloman SA. 2011. The meaning of cause and prevent: the role of causal mechanism. *Mind Lang.* 26(1):21–52
- Wegner DM. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press
- Weidenfeld A, Oberauer K, Hörnig R. 2005. Causal and noncausal conditionals: an integrated model of interpretation and reasoning. *Q. J. Exp. Psychol. A* 58(8):1479–513
- Weiner B. 1995. *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. New York: Guilford
- Wells GL, Gavanski I. 1989. Mental simulation of causality. *J. Personal. Soc. Psychol.* 56(2):161–69
- White PA. 2006. The causal asymmetry. *Psychol. Rev.* 113(1):132–47
- White PA. 2013. Singular clues to causality and their use in human causal judgment. *Cogn. Sci.* 38:38–75
- Wiegmann A, Waldmann MR. 2014. Transfer effects between moral dilemmas: a causal model theory. *Cognition* 131(1):28–43

- Wolff P. 2007. Representing causation. *J. Exp. Psychol.: Gen.* 136(1):82–111
- Wolff P. 2013. Representing verbs with force vectors. *Theor. Linguist.* 38:237–48
- Wolff P, Song G. 2003. Models of causation and the semantics of causal verbs. *Cogn. Psychol.* 47:276–332
- Woodward J. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford Univ. Press
- Zultan R, Gerstenberg T, Lagnado DA. 2012. Finding fault: counterfactuals and causality in group attributions. *Cognition* 125(3):429–40