Psychological Bulletin
2014, Vol. 140, No. 1, 109–139

0033-2909/14/$12.00 DOI: 10.1037/a0031903

Reasoning About Causal Relationships: Inferences on Causal Networks

Benjamin Margolin Rottman
The University of Chicago

Reid Hastie
The University of Chicago Booth School of Business

Over the last decade, a normative framework for making causal inferences, Bayesian Probabilistic Causal Networks, has come to dominate psychological studies of inference based on causal relationships. The following causal networks—[$X \rightarrow Y \rightarrow Z$, $X \leftarrow Y \rightarrow Z$, $X \rightarrow Y \leftarrow Z$]—supply answers for questions like, "Suppose both X and Y occur, what is the probability Z occurs?" or "Suppose you intervene and make Y occur, what is the probability Z occurs?" In this review, we provide a tutorial for how normatively to calculate these inferences. Then, we systematically detail the results of behavioral studies comparing human qualitative and quantitative judgments to the normative calculations for many network structures and for several types of inferences on those networks. Overall, when the normative calculations imply that an inference should increase, judgments usually go up; when calculations imply a decrease, judgments usually go down. However, 2 systematic deviations appear. First, people's inferences violate the Markov assumption. For example, when inferring Z from the structure $X \rightarrow Y \rightarrow Z$, people think that X is relevant even when Y completely mediates the relationship between X and Z. Second, even when people's inferences are directionally consistent with the normative calculations, they are often not as sensitive to the parameters and the structure of the network as they should be. We conclude with a discussion of productive directions for future research.

Keywords: causal inference, causal structures, Bayes nets, Markov assumption, discounting

Most human judgments under uncertainty involve reasoning about causal relationships. For example, a physician tries to infer which disease is the most likely cause of a patient's symptoms (effects). Then, the physician intervenes to alleviate the symptoms by changing the causal dynamics within the patient. Or a corn futures trader forecasts the price of corn by considering the consequences of various possible economic and geopolitical events (e.g., Will a change in China's trade policy influence the value of corn in North America?). And, more personally, one commits to an exercise and diet plan because one believes that the program will produce specific health benefits.

However, until recently, the role of causal reasoning in judgments under uncertainty has been neglected in psychological research. One reason for this neglect has been the lack of a good normative model for the reasoning process that underlies even simple everyday causal inferences, such as in the examples above. In the past 10 years, there has been a paradigm shift in behavioral research on causal inference. The shift has been driven by the dissemination of the Bayesian Probabilistic Causal Network approach to modeling causality (henceforth referred to as "causal networks"). This approach provides prescriptions for rational calculations for inferences on causal networks. The approach has its roots in theoretical articles by Pearl (1988, 2000), Lauritzen and Spiegelhalter (1988), and Spirtes, Glymour, and Scheines (1993) in mathematics and statistics. It has been communicated to behavioral scientists in books by Glymour (2001) and Sloman (2005), as well as articles by many other researchers (Danks, 2009; Gopnik et al., 2004; Rehder & Hastie, 2001; Steyvers et al., 2003; Waldmann, 1996; Waldmann & Martignon, 1998).

Our focus here is on deliberate and partly conscious reasoning about causal beliefs. For example, when our car fails to start one morning, we engage in a deliberate, partly verbalizable sequence of inferences based on our beliefs about what is causing what within the car and its immediate environment: Could something about the weather—recent rainfall—have interfered with the normal sequence of events that occur after we turn the ignition key? Or could the gas tank be empty, the battery be dead, a fuse blown, or a wire chewed through by a squirrel? Here we start from a single fact or set of facts (the car won't start, and it rained last night) and then reason within a system of beliefs (about how the car works) to update our beliefs about the world (rain probably caused a short). Our focus here is not on how we obtain knowledge about how the car works but rather on how we make inferences or judgments about the car given our knowledge of how the car works.

Introduction to Causal Networks

Throughout this article we refer to a stylized example about farming represented in Figure 1. Imagine a farmer who grows cantaloupes and tomatoes. Both cantaloupes and tomatoes are damaged by an early frost (F); they are effects of a common cause.

This article was published Online First April 1, 2013.

Benjamin Margolin Rottman, Section of Hospital Medicine, Department of Medicine, The University of Chicago; Reid Hastie, The University of Chicago Booth School of Business.

This research was supported by National Institutes of Health Grant 1F32HL108711 and The University of Chicago Booth School of Business.

Correspondence concerning this article should be addressed to Benjamin Margolin Rottman, Section of Hospital Medicine, Department of Medicine, The University of Chicago, 5841 South Maryland Avenue, MC5000, Chicago, IL 60637. E-mail: benjaminrottman@uchicago.edu

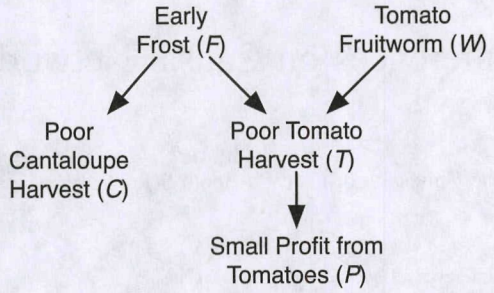


Figure 1. Farming scenario.

In addition, the tomato harvest (T) is hurt by the tomato fruitworm (W); however, this pest does not affect the cantaloupe harvest (C). Finally, if a farmer has a poor tomato harvest, then he or she is likely to reap a small profit from the tomatoes (P).

The graph in Figure 1 conveys the structure of the causal relationships. The nodes represent variables that can take on multiple values. For example, uppercase F represents whether there was an early frost or not. Lowercase represents the state of the node; $f = 1$ denotes that there was an early frost, and $f = 0$ denotes that there was not. The causal relationships are represented by arrows (or "edges") between the nodes.

A fully realized causal network also contains parameters. "Base rate" parameters capture the probability of exogenous nodes, P(F) and P(W), which do not have any explicitly represented causes. "Strength" parameters model how likely each cause is to generate or inhibit each of its effects. When multiple causes influence the same effect (such as F and W on T), a function must be identified to describe how these causes combine to produce the effect.

Once we know the structure and parameters, the normative theory of graphical causal models prescribes how one should infer the state of one variable given the state of another. For example, suppose we learn that a farm had a tomato fruitworm infestation. We would infer that the farm probably had a poor tomato harvest, but we would not rationally infer anything about the cantaloupe harvest. This sort of inference is often called an inference from an *observation*; we observe the state of one variable and then infer another. We also discuss inferences from *interventions*, when we manipulate the state of one variable and then infer the state of another (e.g., if we spray the tomatoes with a pesticide to prevent a fruitworm infestation and then infer the tomato harvest). We also consider reasoning about counterfactuals such as "What would the profit from tomatoes have been had the tomato fruitworm infestation not occurred?" All of these questions can be interpreted as inferences on the causal network in Figure 1.

So, what makes a graph and parameters of this type a *causal* network rather than merely a "probability graph"? It is simply the interpretation of the graph. If the graph is defined as representing causal relationships, then it's a causal graph. However, certain conventions of these networks convey a distinctly causal interpretation. These include (a) the interpretation of arrows as indicating temporal ordering on the variables, (b) the assumption that interventions on the value of one node will be propagated only "downstream" to future states of other nodes, and (c) the presumption that counterfactual inferences can be made about "what would have happened" if the states of nodes had been otherwise than what they in fact were. In sum, the *causal* interpretation of these graphs comes from how they are used and what they represent, not from the probability calculus itself.

Three Steps of Causal Inference

To clarify our focus, we distinguish *three steps of causal inference*: (a) learning the structure of the causal network, (b) learning the parameters, and (c) making a judgment about one node given our knowledge about the other nodes. Our review focuses on making judgments. However, all of the behavioral experiments "teach" the structure and parameters to the human research participants in some manner. Because learning the structure and parameters conceptually precedes making judgments, we provide a brief overview of these prior types of learning.

Learning the structure of a causal network (i.e., which variables cause which other variables) often occurs through explicit teaching (e.g., in a biology class or reading *The Economist*) and deducing plausible causal pathways based on mechanistic hypotheses (e.g., rain water could have caused a short in the car ignition system). Additionally, much of the recent research on "causal learning" has focused on how people learn causal structures from experience (e.g., Gopnik et al., 2004; Lagnado & Sloman, 2004, 2006; Rottman & Keil, 2012; Steyvers et al., 2003; see Lagnado, Waldmann, Hagmayer, & Sloman, 2007, for a summary). For example, a parent might form beliefs about how to raise a well-behaved child by observing *correlations* between children's behaviors and the behaviors of those children's parents. Of course, it is notoriously difficult to learn *causal* relationships from correlations alone. A second way to learn causal structures is from "interventions": A parent might try various child-rearing habits to see which one works best. Finally, people also learn causal structures from a variety of temporal cues. For example, a mother might infer different causal relationships if she notices that after her son has a restless night he misbehaves versus the observation that after he misbehaves he sleeps poorly.

It is still unclear how successful we are at learning causal structures from experience. Furthermore, we often have beliefs about what causes what (e.g., rain might have caused an ignition short in my car) and can make judgments and decisions (e.g., I'll wait to see if the short is fixed after the water dries) without having to learn the causal structure through some form of statistical induction from experience.

The second step is learning the causal strengths, that is, the degree to which a cause influences each of its effects (see Hattori & Oaksford, 2007, for a summary of 41 potential models). Studies investigating learning focus on scenarios when there are one or more possible causes (A, B, C) of a single effect (E), and the goal is to learn the strengths of the alternate causes. Often these experiments do not distinguish whether participants learned about whether the link $A \rightarrow E$ exists versus the strength of the link $A \rightarrow E$; thus, experiments about "causal strength learning" and "causal structure learning" as well as "multiple cue learning" and even "covariation detection" can overlap.

Earlier literature on causal strength learning often focused on "irrational" inferences like illusory correlation (e.g., Jenkins & Ward, 1965) and how strengths could be learned through associative mechanisms (e.g., Dickinson, Shanks, & Evenden, 1984). However, the recent trend has been to focus on rational explana-

tions for patterns in causal strength learning such as conditioning on alternative causes (e.g., Spellman, 1996; Waldmann, 1996; Waldmann & Hagmayer, 2001), accounting for ceiling and floor effects (Cheng, 1997; Novick & Cheng, 2004), understanding the interaction between whether a link exists and the strength of the link (Griffiths & Tenenbaum, 2005), and incorporating prior beliefs about the likely strength of potential causes (Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008).

Regarding the current review, most of the studies that focus on judgment have simply told experimental participants the causal structure rather than having them learn it from experience. Participants learned the parameters by observing correlations between causes and effects or from textual descriptions or prior knowledge (sometimes participants did not have any specific quantitative knowledge of the parameters). Thus, even in controlled experiments there may be questions about participants' beliefs about the causal system (we attend to these issues on a study-by-study basis in this review). Overall, our focus is on people's judgments given their causal belief system.

Simplifications and Limitations of Causal Networks

It is important to keep the nature of the simplifications that are inherent in the causal network framework clearly in mind, as this approach to human judgment depends upon accepting that such simplifications do not drastically distort everyday habits of thinking about causal relationships. The causal network framework is very flexible and can be extended to loosen various assumptions. However, the standard framework—the one that has been the primary focus in the causal reasoning literature—makes the following assumptions.

First, the networks we review do *not* represent any temporal durations such as the length of delay between a punctate cause and effect or the timing of a maximum effect (e.g., ibuprofen has its maximum effect at about 1 hr after ingestion). We note, however that standard causal networks can be expanded to include temporal information (e.g., Buchanan & Sobel, 2011; Rottman & Keil, 2012).

Second, the causal networks we review are acyclic; they cannot have any loops like $X \rightarrow Y \rightarrow Z \rightarrow X$ or "bidirectional" relationships like $X \leftrightarrow Y$. In acyclic networks each variable can be represented as a function of the variables that directly cause it, but if a variable causes itself then this function is indeterminant. Standard networks can be "unfolded over time" to account for causal loops (e.g., Griffiths & Tenenbaum, 2009; Kim, Luhmann, Pierce, & Ryan, 2009; Rehder & Martin, 2011).

Third, the networks considered in most applications are incomplete. Surely there are many variables that could be added that precede, mediate, and/or follow the variables explicitly represented in any network (e.g., other causes and effects of tomato fruitworms or small profits).

Fourth, there are many "zero links" in the network, when in reality there are small causal influences between relevant causal events. For example, in a realistic economic context, the cantaloupe harvest probably has an impact on the market price of tomatoes, but this influence is ignored in the Farming Scenario. This sparseness is also typical of all the relevant behavioral research.

Fifth, an essential property of causal networks is the Markov Assumption. In reference to Figure 1, this assumption says that when the state of T is known, and we infer P, F does not provide any additional information about P. In other words, T completely mediates the relationship from F to P. The Markov assumption greatly simplifies normative causal inference because it identifies variables that can be ignored for certain inferences. The Markov assumption cannot be relaxed or abandoned.

These limitations have led some philosophers and mathematicians to conclude that the entire enterprise of modeling realistic situations with such graphs is futile (e.g., Cartwright, 1999, 2001, 2002; see also articles in Gelman & Meng, 2004). We still believe that the approach helps us understand real causal systems and how ordinary people think about causality. But not all readers will agree, and we want to be clear about the strong assumptions required to believe that Causal Networks provide a useful tool for understanding causal cognition.

Simplifications and Limitations of Psychological Research on Causal Networks

In addition to the limitations and simplifications of the normative causal network approach, there are additional simplifications in the ways that causal inference is typically studied in psychology experiments. First, although the variables in the example network could be continuous, ordinal, or categorical, the majority of behavioral research has focused on binary causes and effects (e.g., the tomato harvest was good or poor, not number of tons of tomatoes harvested). Second, although each causal relationship could be generative or inhibitory, most of the existing research has focused on generative links. In the farming example we represented Early Frost as causing a Poor Tomato Harvest, not preventing a Good Tomato Harvest.

Third, when two or more causes influence one effect, the causes can potentially combine in many different ways. For example, when causes are multivalued, they could produce the effect additively, multiplicatively, or with any other function (Waldmann, 2007). However, most research (which has focused on independent, generative, binary causes) has assumed a particular "functional form" called the "Noisy-OR gate" (e.g., Cheng, 1997; Griffiths & Tenenbaum, 2005; Novick & Cheng, 2004; Pearl, 1988; see Yuille & Lu, 2008, for other functional forms). For example, one might believe that the probability of a poor tomato harvest is determined by the union (as opposed to the intersection, or some other function) of a frost *or* an infestation that successfully causes a poor tomato harvest.

Plan for This Review

Our focus is on how people make inferences and whether their inferences agree with the normative calculations on causal networks. We first discuss whether people's inferences follow the Markov Assumption, which simplifies reasoning by identifying which nodes are relevant for making a particular inference. The rest of the article focuses on how people make use of the parameters of the causal structures. We look at whether people's inferences go in the predicted directions, as well as how close people's inferences come to the normative calculations. We analyze these questions for a variety of different types of paradigmatic causal

structures including chains, common cause structures, one-link structures, common effect structures, and diamond structures.

We finish by making some observations about the quality of human reasoning about causal relationships. To foreshadow our conclusions, many aspects of human reasoning about causal systems reflect the qualitative prescriptions of the normative model. When the calculations imply the probability of an event should increase, usually judgments go up; when they imply a decrease, they go down. But, there are some reliable anomalies. In particular, people seem not to respect the Markov Assumption, and their inferences tend to be weaker than would be implied by the normative model. We also comment on the value of comparing behavioral results to a normative model. Among other reasons, we submit that the comparison is useful because it identifies potential pitfalls for human reasoning about practical matters.

The Markov Assumption

The Markov Assumption identifies which nodes are relevant to an inference and which nodes are irrelevant. Consider the chain in Figure 2, which is a subgraph from Figure 1. Suppose that we are trying to infer whether there will be a large or small profit from tomatoes this year. If we know that there was an early frost, we would be likely to infer a poor tomato harvest and thus a small profit. The probability of $p = 1$ is higher given that $f = 1$ than given $f = 0$; $P(p = 1|f = 1) > P(p = 1|f = 0)$.

However, suppose that we already know that there was a poor tomato harvest ($t = 1$) and later learn that it was caused by an early frost ($f = 1$). Given the poor tomato harvest we would already have inferred that there is likely to be a small profit from tomatoes, and learning that there was or was not an early frost does not change the inference about the tomato profit: $P(p = 1|t = 1) = P(p = 1|t = 1, f = 1) = P(p = 1|t = 1, f = 0)$. F is irrelevant to P once T is known. The technical term for this relationship is that early frost and small profit from tomatoes are "d-separated" by poor tomato harvest; profit is no longer *dependent* on frost once the mediator (poor harvest) is known. These inference patterns are symmetric. For $F \rightarrow T \rightarrow P$, once T is known, learning the state of P does not affect the inference of F.

More generally, the Markov Assumption states that a given node, conditional on all its direct causes, is statistically independent of all other nodes that are not its direct or indirect effects. (See Charniak, 1991, and Sloman, 2005, for gentle introductions to causal graphical models, and Jensen & Nielsen, 2007, for a more technical introduction.) The Markov Assumption becomes even more useful in structures with large numbers of variables because the Markov Assumption may be able to label many of them as irrelevant for a given inference.

The common cause graph works much like the chain. If we find out that there was a poor tomato harvest, we might infer that there was an early frost and thus that there was also a poor cantaloupe harvest. However, if we already know that there was an early frost, then we would predict that there was a poor cantaloupe harvest regardless of whether there was a poor tomato harvest or not.

For the common effect structure, neither F nor W have any direct causes in the network, so they are unconditionally independent. Just because there was an early frost does not mean that there was a fruitworm infestation, or vice versa. (In some modeling applications exogenous causes like F and W are not necessarily assumed to be independent.)

Evidence of the Use of the Markov Assumption for Inferences

The Markov Assumption identifies which variables can be ignored for particular inferences, simplifying the inference process. Rehder and Burnett (2005) provided the first comprehensive test of the Markov Assumption. Here is an example of one scenario they used involving a causal chain. Participants learned about Kehoe ants, which typically have blood high in *i*ron sulfate, which causes a hyperactive immune *s*ystem, which causes *t*hick blood, which causes them to build nests *q*uickly [$I \rightarrow S \rightarrow T \rightarrow Q$], but participants were not given the specific parameters of the causal model. Participants were then presented with an ant with certain features such as [$s = 1, t = 1, q = 0$] and were asked to infer the probability of I. Whether T and Q are 1 or 0 should not affect the inference of I because S is known to be 1.

Rehder and Burnett (2005) found that participants systematically violated the Markov Assumption. Throughout the article, we use C to refer to an exogenous cause (a factor that does not have any known causes in the network), M to refer to a mediator, and E to refer to an effect (a node that does not cause any other nodes in the network). For the chain structure (see Figure 3), even when they knew the state of M_1, if M_2 and E were present, then they were more likely to infer that C was present. There were analogous effects for inferring E. For the common cause, even if they knew

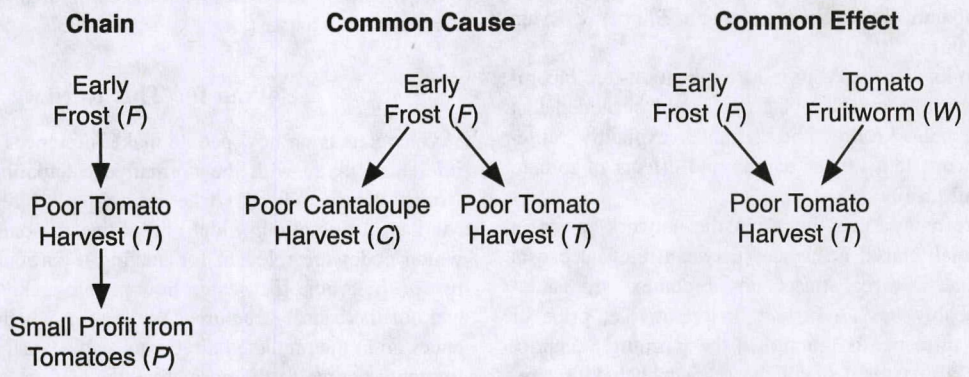


Figure 2. Three prototype causal networks embedded in the farming scenario.

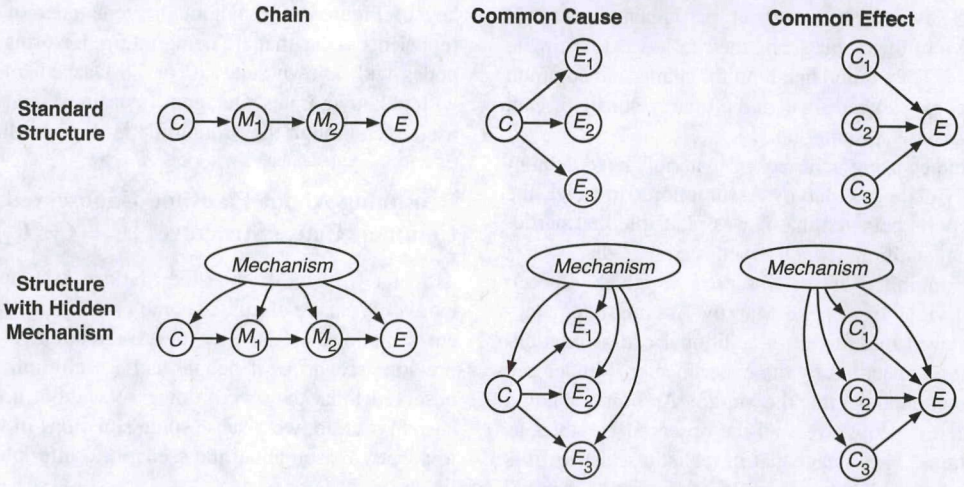


Figure 3. Causal structures investigated by Rehder and Burnett (2005). C = an exogenous cause (a factor that does not have any known causes in the network); M = a mediator; E = an effect (a node that does not cause any other nodes in the network).

the state of C, the states of E_2 and E_3 influenced participants' inferences of E_1. For the common effect structure, if C_2 and C_3 were present (and the state of E was unknown), participants were more likely to infer that C_1 was present.

In order to account for these violations of the Markov Assumption, Rehder and Burnett (2005) suggested that their participants inferred that there was another feature, an unobserved "mechanism" that was a direct cause of all other features, somewhat like a category essence (see Figure 3, bottom row). With the unobserved mechanisms, all the features that were previously independent are dependent because they are common effects of the unobserved mechanism. For example, in the chain structure, even when the state of M_1 is known, if M_2 is present then the mechanism is more likely to be present, and thus C is more likely to be present.

Explaining this violation of the Markov rule by assuming participants had "imported" an unobserved cause into their mental representations leaves some open questions. First of all, there is no direct evidence that people believe in this unobserved mechanism. For the case of living kinds, it is plausible to hypothesize factors like DNA that might serve as underlying causes of many causal features. But Rehder and Burnett (2005) also used other categories such as Romanian cars (with features such as butane laden gas and loose fuel filter gaskets). In one experiment they even used "skeletal" categories called "Daxes," and the four features were simply labeled A, B, C, and D with no additional meaning. It is unclear what sort of unobserved mechanism could be posited in these cases. Furthermore, Rehder (2006) replicated these results in scenarios that did not involve categories (e.g., low interest rate → small trade deficit → high retirement savings) as well as with a completely abstract domain (e.g., Variable A → Variable B → Variable C). These experiments suggest that even if the Markov violations can be modeled by adding an unobserved common cause to the structure, it is not obvious why people would assume such a node.

In order to eliminate the unobserved category "mechanism" as a possible explanation for the Markov violations, Rehder (2012) used nodes labeled as causes and effects that were not features of a category (e.g., urbanization causes socioeconomic mobility). Rehder (2012) also wondered if people were inferring other direct causal relationships between the variables based on their prior knowledge, which could lead to apparent violations of the Markov Assumption. Thus, he also counterbalanced the nodes in a way such that systematically inferring additional links between the nodes would not lead to violations of the Markov Assumption. Yet, he still found persistent violations.

Rehder (2012) also tested the effects of deliberative reasoning versus more intuitive judgments. In one condition he required participants to respond to the inference questions in under 10 s, and in another he asked them to justify their inferences. There was no consistent effect of the justification or speeded manipulations; if anything, it appeared that justification led to *more* Markov violations. This pattern of findings suggests that Markov violations are not merely due to a quick intuitive judgment such as associative reasoning.

Burnett (2004) conducted a number of similar experiments and also found significant violations of the Markov Assumption. In addition, he found evidence that people's inferences fit a proximity heuristic: nodes that are closer to the inferred node are weighted more, even when an intermediate node is known. For example, in the chain $C \rightarrow M_1 \rightarrow M_2 \rightarrow E$, when inferring C and the state of M_1 is known, the state of M_2 has a larger impact on C than does the state of E.

Mayrhofer et al. (2010) tested the Markov assumption in a task in which aliens read the minds of other aliens (see the following section for more details about this study). One condition involved a chain $C \rightarrow M_1 \rightarrow M_2 \rightarrow E$, such that Alien E read the mind of Alien M_2, who read the mind of M_1, who read the mind of C. Participants inferred that Alien E's thoughts were almost entirely dependent upon M_2 (and very weakly dependent upon C and M_1). This particular domain and chain structure (essentially the "telephone game") seems to emphasize to participants that only the direct cause is relevant for any given inference.

Finally, Sussman and Oppenheimer (2011) conducted a study in which they told participants causal relationships between three

fictitious plumbing devices. On each trial, participants were told integer values of two of the devices, and their task was to estimate the value of the third. They found that both for chains and common cause structures, people showed small and probably nonsignificant violations of the Markov Assumption.

In sum, many studies using a variety of materials have demonstrated that people violate the Markov Assumption. However, the authors of these reports believed that it was plausible that participants imagined an additional unobserved variable that was a common cause or inhibitor of the observed variables. Burnett (2004) even called violations of the Markov Assumption "adaptive" if people believed that there are additional causal relationships aside from those specified by the experimenter. Rehder and Burnett (2005) also pointed out that the Markov Assumption could appear to be violated if people treat all the observed variables as imperfect observations. This means that in realistic scenarios it is very difficult to rule out all rational explanations for "apparent" violations of the Markov Assumption.

At the same time, a number of studies have used scenarios in which there is no compelling reason why people would infer additional causal links. It is also notable that the Markov violations always seem to be "positive." For $A \rightarrow B \rightarrow C$, people essentially infer an additional positive correlation between A and C above and beyond the correlation implied by B. If people were really inferring additional unobserved links, it is unclear why these links would overwhelmingly be positive. Thus, some of these inferences seem to be true violations of the Markov Assumption in that there is no plausible adaptive reason for inferring an unobserved common cause given the particular cover story. We summarize the results of this section in Figure 4 and use the same notation presented in the key of Figure 4 throughout the remainder of this review. Bold represents nodes that are being inferred. Normal weight represents nodes with known states (0 or 1). Dashed lines represent nodes with unknown states. Octagons (stop sign) represent nodes that are used even though they should be ignored for the given inference.

Reasoning About Plausible Unobserved Links on Common Cause Structures [$E_1 \leftarrow C \rightarrow E_2$]

So far we have framed the Markov Assumption as being normative. We have discussed some potential explanations for apparent violations of the Markov Assumption. However, in all the previous scenarios, if people had actually inferred additional unobserved links they were doing so without good reasons. In the current section, we discuss some situations in which people seem to adeptly reason about the scenario to infer plausible unobserved links.

In a standard common cause structure, $E_1 \leftarrow C \rightarrow E_2$, we can conceive of the two effects as having additional *independent* unobserved influences (see the *U*s in Figure 5a). Though Figure 5a is the standard way to interpret common effect structures when we have no additional information, there are some situations in which we might believe that the effects would be correlated above and beyond what would be implied by C alone. Figures 5b and c represent two such structures (see also the Feature Uncertainty Model in Rehder & Burnett, 2005).

Mayrhofer et al. (2010) investigated a social transmission scenario and found that describing the nodes as either active or passive moderated whether people interpreted the structure as having independent versus correlated errors. They used a cover

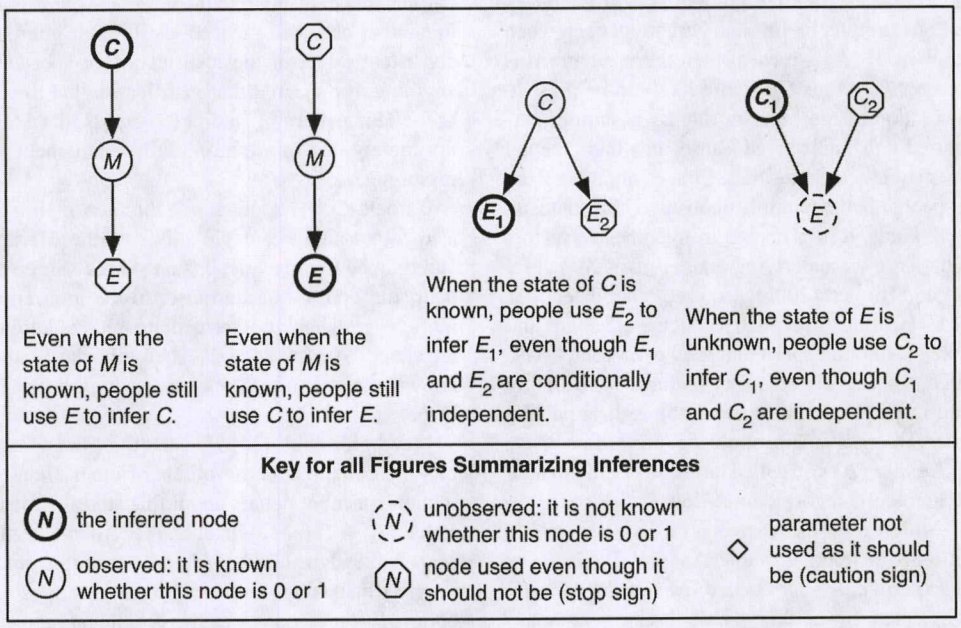


Figure 4. Summary of Markov assumption violations. Bold represents nodes that are being inferred. Normal weight represents nodes with known states (0 or 1). Dashed lines represent nodes with unknown states. Octagons (stop sign) represent nodes that are used even though they should be ignored for the given inference. C = an exogenous cause (a factor that does not have any known causes in the network); M = a mediator; E = an effect (a node that does not cause any other nodes in the network).

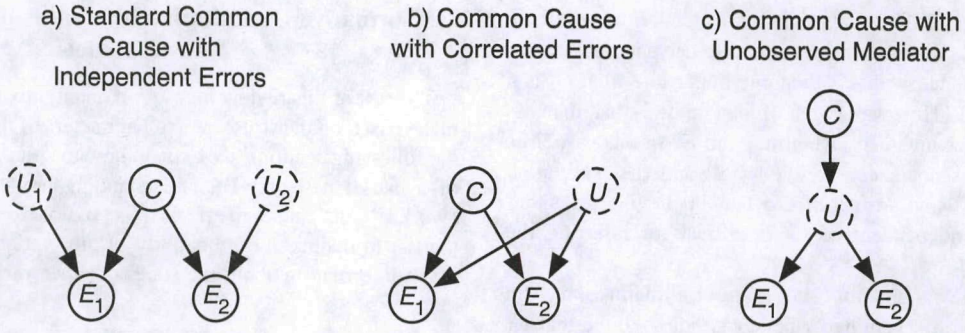


Figure 5. Independent versus correlated errors on a common cause network. Dashed U nodes represent unobserved influences.

story about four telepathic aliens who could transfer their thoughts (either "por" = 1 or "tus" = 0) through mind reading. In the "active" condition, one alien (cause; C) was described as *sending* his thoughts to the effect aliens (Es). Participants inferred $P(e_2 = 1|c = 1, e_1 = 1) > P(e_2 = 1|c = 1, e_1 = 0)$. When C was described as the agent "sending" the message to E_1 and E_2, one could plausibly reason that something could cause an error in the transmission of the message to *both* E_1 and E_2 (*e.g.,* Alien C wasn't concentrating hard enough). This "sending" condition seems to imply correlated errors (e.g., Figure 5b or c). In contrast, when the effect aliens were described as passively "reading" the mind of Alien C, there was a smaller difference between $P(e_2 = 1|c = 1, e_1 = 1)$ and $P(e_2 = 1|c = 1, e_1 = 0)$. A plausible reason is that if one effect alien misread the message, it should not have an impact on another alien's ability to read the message, with independent errors like as depicted in Figure 5a.

Mayrhofer, Goodman, Waldmann, and Tenenbaum (2008) investigated another aspect of a causal scenario likely to convey beliefs in unobserved correlated errors. They used the same alien cover story, but now there were two different types of effect aliens, green and yellow. When one yellow alien misread the message people tended to infer that another yellow alien would also misread the message, but whether the green aliens correctly read the message did not matter for inferring whether a yellow alien would correctly read the message. This pattern can be interpreted as indirect evidence for two sets of correlated errors for the two types of aliens.

Walsh and Sloman (2004, 2007) also investigated rational explanations for correlations between effects of a common cause above and beyond the correlation implied by the cause. They used realistic common cause scenarios (e.g., jogging causes increased fitness level and weight loss). When told that Tim did not lose weight, people often came up with explanations that were common causes or disablers of both effects (e.g., jogging increased Tim's appetite, which caused him not to lose weight and prevented his fitness level from increasing).

In sum, the studies in this section have identified a number of scenarios for which it seems reasonable for people to use their own prior knowledge or information conveyed in the description of the scenario to infer structures with correlated errors. However, the fact that inferring such correlated errors was to be expected in these studies does not diminish the fact that in the studies in the previous section there was no similar reason to infer correlated errors and thus no compelling reason to believe that the Markov Assumption did not hold.

Beliefs About Whether the Causes in a Common Effect Structure [$C_1 \rightarrow E \leftarrow C_2$] Are Correlated

For common effect structures, $C_1 \rightarrow E \leftarrow C_2$, the strict interpretation of the Markov Assumption implies that the two exogenous causes, C_1 and C_2, are independent from each other. Yet in practice, statistical causal modelers (e.g., LISREL) often allow for the possibility that they are correlated. For example, imagine an economist considering three economic indexes postulated to have the following structure: $I_1 \rightarrow I_3 \leftarrow I_2$. The modeler would likely want to test for the possibility that I_1 and I_2 are correlated rather than just assume that they are independent. In this section we discuss situations in which people believe that C_1 and C_2 are correlated, exploring how these beliefs are influenced by different types of experience and whether people's beliefs are consistent.

As already mentioned, Rehder and Burnett (2005) told participants a cover story involving a common effect structure $C_1 \rightarrow E \leftarrow C_2$. The participants treated C_1 and C_2 as correlated even though there was no obvious compelling reason to do so. Von Sydow, Hagmayer, Meder, and Waldmann (2010; Experiment 2) told participants the structure $C_1 \rightarrow E \leftarrow C_2$. In a set of learning trials participants observed whether each variable was present or absent; C_1 and C_2 were uncorrelated. Afterward, participants inferred that C_1 and C_2 were independent, $P(c_2 = 1|c_1 = 1) = P(c_2 = 1|c_1 = 0)$. Thus, even if people tend to believe that C_1 and C_2 are correlated, they can fairly quickly learn that that C_1 and C_2 are independent.

Hagmayer and Waldmann (2000) conducted a similar study. On a given learning trial, however, participants saw either C_1 and E or C_2 and E, so they could not calculate the correlation between C_1 and C_2. At the end of the learning trials, participants judged $P(c_2 = 1|c_1 = 1)$ and $P(c_2 = 1|c_1 = 0)$, which were converted into the correlation measure *phi*. The correlations in two experiments were slightly positive (.16 and .24). Perales et al. (2004) conducted a parallel study. In most conditions, participants inferred correlations close to zero, but in one condition with strong causal relationships about one third of the participants inferred a substantial positive correlation.

The assumption that C_1 and C_2 are independent is particularly important for inferring the causal strength of C_1 on E *when* C_2 *is*

unobserved (Cheng, 1997). Suppose that C_1 and E are strongly correlated. If one believes that there are no other potential causes of E that are correlated with C_1, then one might infer that C_1 is a strong cause of E. However, what if one knows that there is another factor, C_2, which causes both C_1 and E? In this case, it is possible that C_1 is not a cause of E at all and the correlation between C_1 and E is an artifact of C_2. Thus, believing that other causes of E are independent of C_1 is critical for inferring the strength of C_1.

Hagmayer and Waldmann (2007) and Luhmann and Ahn (2007; Experiment 3) examined whether people believe that C_1 is independent from an *unobserved* C_2. On each trial people observed C_1 and inferred whether C_2 was present or absent. Both of these studies found that people judged C_1 and C_2 to be correlated. More importantly, the estimated correlation depended on the learning conditions. In Hagmayer and Waldmann's (2007) Experiment 1, when the two causes were relatively weak, people thought that C_1 and C_2 were positively correlated; when they were relatively strong, people thought they were negatively correlated. But Luhmann and Ahn (Experiment 3) found a different result: When C_1 had a positive influence on E, people inferred a positive correlation, but in the condition in which C_1 had zero effect on E, people inferred a negative correlation. These results are surprising because there is no normative reason why changing the strengths should lead people to change their inferences of the correlation between C_1 and C_2. Hagmayer and Waldmann (2007) also asked people to make summary judgments of $P(c_2 = 1|c_1 = 1)$ and $P(c_2 = 1|c_1 = 0)$ at the end of the learning trials. Unlike the trial-by-trial judgments, these judgments reflected a belief that C_1 and C_2 were *independent*. It is surprising and unclear why these judgments were inconsistent.

Until now we hedged about what people should infer about the correlation between C_1 and C_2, proposing that people's inferences should merely be consistent. However, Hagmayer and Waldmann (2007, Experiment 2) conducted a study that normatively implies that C_1 and C_2 are independent. On each trial participants *chose* whether C_1 occurred or not and then inferred whether C_2 would be present or not. Because participants chose C_1 without knowing C_2 or E_2, this intervention should be interpreted as cutting any links to possible unobserved common causes. Yet, participants usually inferred that C_1 and C_2 were negatively correlated. In sum, people's beliefs about the relationship between C_1 and C_2 are inconsistent and in one instance go against the normative framework.

Summary

The Markov Assumption greatly simplifies learning and reasoning with causal networks. However, people appear to be unaware of the simplicity it affords. When making inferences, people often use nodes that, according to the Markov Assumption, are irrelevant for the particular inference. Furthermore, related research outside the focus of this review also shows that people fail to capitalize on the Markov Assumption when learning causal networks (Steyvers et al., 2003; Experiment 3; Fernbach & Sloman, 2009; Jara, Vila, & Maldonado, 2006).

Normative Quantitative Inferences on Graphical Causal Models

The rest of this review focuses on quantitative inferences people make based on the structure and parameters of the causal model. In the following sections, we explain how to simulate the functioning of a causal network. By understanding how a causal structure "works," from causes to effects, it is possible to make inferences—that is, to deduce the probability of any variable in the network given information about the states of other variables.

Parameterizing a Structure: Modeling How Causes Combine to Produce an Effect

The first task required to make quantitative inferences on a causal network is to model how each individual node is produced by its direct causes, otherwise known as the "parameterization" of the model. We start with a one-link structure $C \rightarrow E$. One common way to conceive of $C \rightarrow E$ is with an additional alternative unobserved cause of E, which we call A; $C \rightarrow E \leftarrow A$. A represents the "causal background" or the likelihood of other possible factors that we cannot directly observe generating E (and they are assumed to be independent of C). A psychological explanation for adding A into the model is that if E ever occurs without C, then we must believe that some other cause produced E. We denote the likelihood of A (a = 1) generating E ($e = 1$), or the "strength" of A as S_A, which equals $P(e = 1|c = 0)$. (Metaphysically, S_A reflects both the probabilities of all the unobserved causes as well as the strengths of all these unobserved causes. But because we do not know specifics about the probabilities and strengths of all these unobserved causes, we use S_A for simplicity.) The two other parameters of the model are the base rate of C, $P(c = 1)$, and the strength of C causing E, S_C. C can cause E only when C is present.

According to this parameterization, E can be produced two ways: C can produce E, with the probability $[P(c = 1)S_C]$, and A can produce E with the probability S_A. Thus, $[1 - P(c = 1)S_C]$ is the probability that C fails to generate E, and $[1 - S_A]$ is the probability that A fails to generate E. Because we are assuming that C and A are independent, the probability of E occurring is 1 − the probability that C and A both fail to produce E (the probabilistic *union* of either C or A generating E, see Table 1 row 1). This is the Noisy-OR combination rule for two independent causes (see Cheng, 1997; Pearl, 1988).

This same logic can be extended to cases with two or more generative causes, all of which can independently produce E (Table 1 row 2). To determine the union of any of the three causes successfully producing E, one can calculate 1 minus the probability of all of the generative causes failing to produce $e = 1$.

What if C inhibits or decreases the probability of E on a $C \rightarrow E$ structure? The standard function to represent an inhibitory cause is called "noisy-And-Not." A can produce $e = 1$ with the probability S_A. S_C is the probability that C would inhibit $e = 1$, so $(1 - S_C)$ is the probability that C fails to inhibit $e = 1$. Thus, $P(e = 1)$ is the product of A generating E, and C failing to inhibit E (row 3 in Table 1). Rows 4 and 5 show other cases that can be determined with the same logic (see Novick & Cheng, 2004).

From the formulas in Table 1 it is trivial to calculate conditional probabilities of E given knowledge of the states of the causes. When a given cause is known to be present (or absent), $P(c = 1)$

Table 1

Probability of an Effect Given Different Combinations of Binary Generative and Inhibitory Causes, Assuming Noisy-Or (Generative) or Noisy-and-Not (Inhibitory) Functions

Structure	$P(e = 1)$
One generative cause plus A	$1 - [1 - S_A][1 - P(c = 1)S_C]$
Two generative causes plus A	$1 - [1 - S_A][1 - P(c_1 = 1)S_{C1}][1 - P(c_2 = 1)S_{C2}]$
One inhibitory cause plus A	$S_A[1 - P(c = 1)S_C]$
Two inhibitory causes plus A	$S_A[1 - P(c_1 = 1)S_{C1}][1 - P(c_2 = 1)S_{C2}]$
One generative cause (C_1) and one inhibitory cause (C_2) plus A	$[1 - [1 - S_A][1 - P(c_1 = 1)S_{C1}]][1 - P(c_2 = 1)S_{C2}]$

simplifies to 1 (or 0). For example, the following two conditional probabilities are deduced from row 1: $P(e = 1|c = 1) = 1 - (1 - S_A)(1 - S_C)$ and $P(e = 1|c = 0) = S_A$. These conditional probabilities will be used in the next section.

There are other ways that multiple binary causes could influence an effect. For example, $P(e = 1)$ could be determined by a simple sum of the strengths of the causes with cutoffs so that $P(e = 1)$ cannot go above 1 or below 0. Or, analogous to a logistic regression, $P(e = 1)$ could be determined by an S-shaped function over the sum of the strengths of the generative and inhibitory causes. Behavioral research has almost exclusively focused on noisy-OR and noisy-AND-NOT functions, so we do not consider these other possibilities any further.

So far we have discussed how to parameterize a structure with multiple causes of a single effect. To parameterize a larger structure, each exogenous node needs a parameter to represent its base rate, and each arrow needs a causal strength parameter. Additionally, if a node ever occurs when its causes are absent, then it also needs an S_A parameter. Figure 6 shows the parameters for five canonical causal structures.

In sum, this section explains how the conditional probability of an effect given its causes can be derived from causal strengths assuming a noisy-OR integration. It is also possible to parameterize a structure at the level of conditional probabilities instead of going down to causal strengths: $C \rightarrow E$ would be parameterized by $P(e = 1|c = 1)$ and $P(e = 1|c = 0)$. Either parameterization works, although reasoning with causal strengths provides a deeper level of analysis and is simpler to represent when there are multiple causes of a single effect.

From Conditional Probabilities to the Factorization and Joint Distribution

The previous section explained how to model the probability of an effect given its direct causes. The second step for representing a causal structure is the joint probability distribution, the probability that the variables in the network are each in a particular state. For the farming example *Early Frost* (F) $\rightarrow$ Poor Tomato Harvest (T), the joint distribution specifies the percentage of farms that experienced an early frost and a poor tomato harvest, $P(f = 1,t = 1)$, the percentage of farms that experienced an early frost but a normal tomato harvest $P(f = 1,t = 0)$, and so on. Determining the joint distribution requires applying the "factorization" of the network. The factorization represents the structure of the graph in terms of conditional probabilities associated with each causal relationship in the graph. For the $C \rightarrow E$ structure, the factorization is simply $P(E,C) = P(E|C)P(C)$. For example, suppose that C is generative and $P(c = .1)$, $S_A = .2$, and $S_C = .5$, and thus $P(e = 1|c = 0) = .2$, and $P(e = 1|c = 1) = .6$. Table 2 shows how to calculate the joint probability distribution for the four joint states of C and E. The four joint probabilities are mutually exclusive and exhaustive, so they sum to 1.

For more complicated causal structures, the factorization of the joint probability distribution works in essentially the same way:

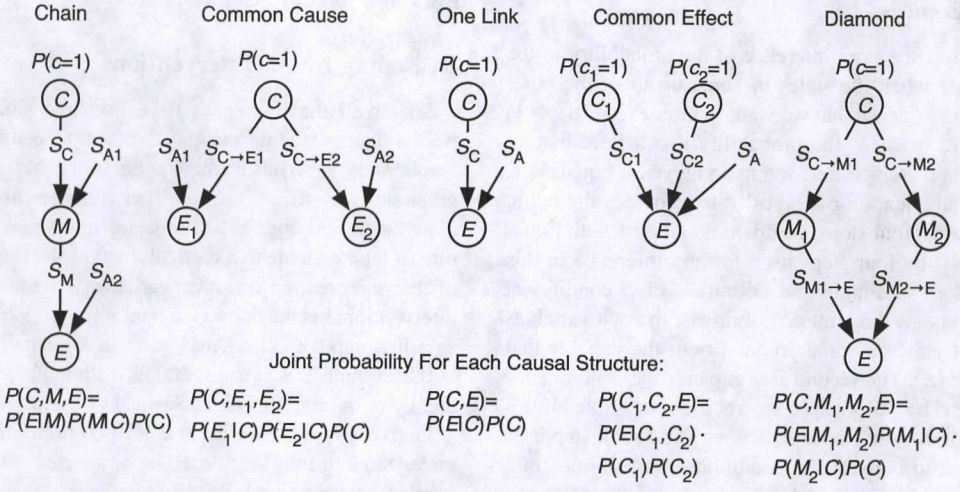


Figure 6. Parameters for five structures.

Table 2
Joint Probability Table for C→E

Joint probability	Factorization
$P(c = 1,e = 1)$	$P(e = 1\|c = 1)P(c = 1) = .6 \times .1 = .06$
$P(c = 1,e = 0)$	$P(e = 0\|c = 1)P(c = 1) = .4 \times .1 = .04$
$P(c = 0,e = 1)$	$P(e = 1\|c = 0)P(c = 0) = .2 \times .9 = .18$
$P(c = 0,e = 0)$	$P(e = 0\|c = 0)P(c = 0) = .8 \times .9 = .72$

The probabilities of each variable given its direct causes are multiplied together (and for exogenous variables with no causes in the network the base rate is used). Figure 6 shows how to calculate the joint probability for five canonical causal structures.

Marginal Probabilities

Whereas a joint probability is the probability of all the nodes in a network assuming a specific set of states, a marginal probability is the probability of a subset of the nodes in the structure assuming a specific set of states. For example, on the $C \rightarrow E$ structure, one might want to know $P(e = 1)$. $P(e = 1)$ can be calculated by summing $P(c = 1,e = 1)$ and $P(c = 0,e = 1)$—that is, rows 1 and 3 in Table 2—which is known as "marginalizing" over C. Note that certain marginal probabilities, such as this one, can also be calculated directly from the parameterization in Table 1. Consider a marginal probability on a structure with three nodes A, B, and C. The marginal probability $P(a = 1,c = 1)$ can be obtained from the sum of the two joint probabilities $P(a = 1,b = 1,c = 1)$ and $P(a = 1,b = 0,c = 1)$, effectively "marginalizing out" B. In sum, marginal probabilities can be calculated by summing over joint probabilities.

Marginal probabilities are important for two reasons. First, they are inferences in their own right. For example, on the chain $C \rightarrow M \rightarrow E$, one might want to know $P(m = 1)$ or $P(e = 1)$. Second, marginal probabilities are important because they are often required when deducing conditional inferences, which is explained in the next section.

From Joint Probabilities and Marginal Probabilities to Conditional Inferences

A conditional inference is an inference of the probability of the state of one variable when the states of some or all of the other variables are known. Suppose that we want to infer $P(c = 1|e = 1)$ on a $C \rightarrow E$ structure (perhaps the probability that a farm had an early frost given that there was a poor tomato harvest). Equation 1, which involves an application of Bayes' rule, provides the math. The derivation requires four steps; compare Equation 1 with Equation 2, which shows the four steps used for any inference in this article. The first step is simply the definition of a conditional probability, which equals the joint probability of the two variables (C and E) divided by the marginal probability of the variable that is conditioned upon (E). The second step expands the denominator by marginalization. The third step converts the joint probabilities into the factorization for the causal structure. The fourth step uses the parameterization to convert the conditional probabilities into the parameters. From the final product, it can be seen that $P(c = 1|e = 1)$ increases as $P(c = 1)$ and S_C increase and as S_A decreases.

This relatively simple math provides the basis for a wide variety of inferences across different types of causal structures. (We make a suggestion when deriving inferences: convert probabilities of the form $P(a = 0)$ to $[1 - P(a = 1)]$ and $P(a = 0|b = 1)$ to $[1 - P(a = 1|b = 1)]$. But remember that $P(a = 1|b = 0) \neq [1 - P(a = 1|b = 1)]$.)

Equation 1. An Inference on a C→E Structure

$$P(c = 1|e = 1) = \frac{P(c = 1, e = 1)}{P(e = 1)} = \frac{P(c = 1, e = 1)}{P(c = 1, e = 1) + P(c = 0, e = 1)}$$
$$= \frac{P(e = 1|c = 1)P(c = 1)}{P(e = 1|c = 1)P(c = 1) + P(e = 1|c = 0)P(c = 0)} = \frac{S_C + S_A - S_C S_A}{S_C + S_A/P(c = 1) - S_C S_A}$$

Equation 2. Canonical Method of Calculating Inferences

$$P(X|Y, Z) = \frac{P(X, Y, Z)}{P(Y, Z)} = \frac{P(X, Y, Z)}{P(x = 1, Y, Z) + P(x = 0, Y, Z)}$$

... convert joint probability into factorization ... convert conditional probabilities into parameters ... simplify

Reasoning Based on Observed Frequencies

So far we have explained how to derive quantitative inferences from the structure and the parameters of the network. However, in many instances in the real world (and in some experiments), people experience the probabilistic relationships between the variables in a network. In such cases participants may rely on memories of specific events rather than reason on the structure itself.

Consider a scenario in which you are told that $C \rightarrow E$, and then you observe whether C and E are present or absent on 20 separate trials. For example, perhaps you observe 20 different farms and note whether each farm had an early frost or not (the cause) and whether each farm had a poor tomato harvest or not (the effect). One could theoretically tabulate the frequencies of C and E to compile Table 2 and perform inferences on Table 2 without using Bayes' rule. For example, $P(c = 1|e = 1) = P(c = 1,e = 1)/P(e = 1)$ = row 1/(row 1 + row 3). However, the number of rows in the joint probability table grows exponentially with the number of variables. The causal network framework greatly simplifies inference because the number of parameters (base rates and strengths) is often much smaller than the number of rows in a joint probability table.

Reasoning About Interventions

So far all the inferences have involved situations in which a person learns about one piece of information and then infers another (e.g., "What is the likelihood of a poor tomato harvest given an early frost?"). Causal networks are also useful for modeling "interventions," when an actor intervenes on a causal structure to set a variable to a particular value and then infers the effects of that intervention on other variables. The ability to distinguish interventional versus observational inferences has often been cited as a hallmark of causal reasoning in humans (e.g., Meder et al., 2008; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005) and even in rats (Blaisdell, Sawa, Leising, & Waldmann, 2006).

Pearl (2000) and Spirtes et al. (1993) presented a framework for understanding interventions. The basic idea is that when an intervention sets a variable to a particular state, it severs all the ties from other causes of the manipulated variable. The intervention

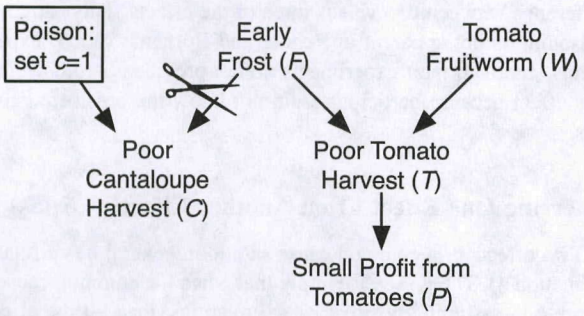


Figure 7. Farming scenario after an intervention poisoning the cantaloupe.

propagates to the effects of the manipulated variable but not to its causes. For example, suppose that a jealous neighbor sprays a poison on the cantaloupes, ensuring a poor cantaloupe harvest. This intervention can be modeled by cutting the link from F to C (see Figure 7). Normally a poor cantaloupe harvest might be a sign that there was an early frost. However, because we know that the cantaloupes were poisoned, we know that there is no longer a relationship between an early frost and the cantaloupe harvest. Once the links to the manipulated variable have been eliminated, all the inferences on the resulting structure are exactly the same as explained above. This method of calculating the effect of interventions is appropriate for "perfect" interventions—when the intervention completely determines the state of the manipulated variable, and the intervention is independent of the rest of the network (Meder, Gerstenberg, Hagmayer, & Waldmann, 2010; Woodward, 2003).

In the next sections, we discuss inferences on various causal structures. Note that the earlier discussion of the Markov Assumption has already noted many inferences on causal structures. Here we discuss the rest of the inferences for which empirical research exists.

Chain $C \rightarrow M \rightarrow E$

Here we discuss transitive and marginal inferences on chains. We skip consideration of inferences about the state of the mediator given C and E because no studies have provided results on the quality of these judgments.

Inferring the Effect From the Cause: Transitive Causal Inferences

Probabilistic causal relations are transitive. On the chain causal structure, if C is known to cause M, and M is known to cause E, then there should be a correlation between C and E. If both links are positive or if both are negative, then the relationship between C and E should be positive. However, if one link is positive and the other is negative, then the relationship between C and E should be negative. Equation 3 shows how to derive the transitive inference. We do not reduce Equation 3 all the way down to causal strengths because the present format in terms of conditional probabilities can be used regardless of whether the links are positive or negative (i.e., $P(e = 1|m = 1) < P(e = 1|m = 0)$).

Equation 3. A Transitive Inference on a Chain

$$P(e = 1|c = 1) = \frac{P(c = 1, e = 1)}{P(c = 1)} = \frac{P(c = 1, m = 1, e = 1) + P(c = 1, m = 0, e = 1)}{P(c = 1)}$$
$$= [P(e = 1|m = 1) - P(e = 1|m = 0)]P(m = 1|c = 1) + P(e = 1|m = 0)$$

Baetu and Baker (2009) had participants learn the contingencies between C and M and between M and E separately and then asked about the relationship between C and E. They found that people generally followed the normative pattern: a positive relation if both links were positive or both were negative, otherwise a negative relation. However, their inferences from C to E were weaker than predicted by Equation 3; that is, the difference between $P(e = 1|c = 1)$ and $P(e = 1|c = 0)$ was too small. Note that participants inferences' were made on a -10 ("when C is 1 it perfectly prevents E from being 1") to $+10$ ("when C is 1 it perfectly causes E to be 1") scale. We describe results on a 0.00 to 1.00 probability scale when we felt that they could be transformed into a probability scale without a significant change in meaning.

Jara et al. (2006) examined the learning of chain structures in a "second-order conditioning" paradigm. Participants saw M paired with E and C paired with M. In one set of experiments, participants inferred that C causes E even though they never saw C and E appear together: They made the transitive inference. In a second set of experiments, after participants learned the $M \rightarrow E$ and $C \rightarrow M$ relationships, they were subsequently presented with a set of trials in which M occurred without E, which was intended to extinguish the $M \rightarrow E$ relationship. Surprisingly, participants still inferred that C would cause E. Some associative models predict that people would form a *direct* association between C and E, contrary to the chain structure.

In another study, participants were told about the chain structure, worked through 192 trials in which they observed C, M, and E, and lastly judged $P(e = 1|c = 1)$ (von Sydow et al., 2010; also see von Sydow, Meder, & Hagmayer, 2009). Normally if the $C \rightarrow M$ and $M \rightarrow E$ links are both positive, then there will be a positive relation from C to E. However, von Sydow et al. (2010) created a set of stimuli in which there was zero

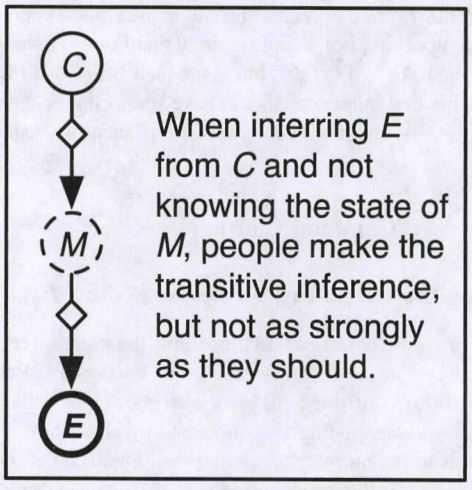


Figure 8. Summary of the $P(E|C)$ inference. See Figure 4 for notational conventions.

correlation between C and E even though the correlations between $C \rightarrow M$ and $M \rightarrow E$ were both positive. Technically their stimuli violated the Markov condition; conditional on M, C and E were negatively correlated. In this way, these experiments were designed to test whether people rely more on the actual observed contingencies or on the transitive relationship implied by a chain structure that is faithful to the Markov Assumption. Even though there was zero correlation between C and E, participants inferred a positive correlation of about .25 (in Experiment 1; .10–.15 in Experiment 2; von Sydow et al., 2010). These results show that people infer transitivity, a conceptual property of causal Bayesian networks, even when the experienced data do not support it.

These three studies suggest that people make transitive inferences from C to E and that these inferences persist even when contradicted by data in which there is no correlation between C and E. But, somewhat paradoxically, when there is a correlation between C and E in the data, the transitive inferences are not as strong as would be predicted by the normative model. One explanation for this pattern of findings is that people's transitive inferences are based on their beliefs about the causal structure (i.e., that transitive inferences are warranted by a chain structure) and are less sensitive to the experienced contingencies. Figure 8 summarizes these findings. Throughout the review, diamonds represent parameters that are not used as they should be (caution sign).

Marginal Probabilities

Rehder and Kim (2010) investigated how people infer the marginal probability of a mediator and effect; $P(m = 1)$ and $P(e = 1)$. They presented people with chain structures and told participants the strengths of the causal relationships. We used participants' inferences of $P(c = 1)$, combined with the strengths that they were given, to model $P(m = 1)$ and $P(e = 1)$.

Overall, participants were sensitive to the qualitative predictions of the normative causal network; however, they were not sensitive enough to the strengths. In one condition (Experiment 2), if a cause occurred its effect would occur 75% of the time, $S_C = S_M = .75$, and the effects would occur only if their causes occurred (i.e., $S_A = 0$). In this case, the marginal probability of each successive node should decrease; however, the decreasing slope was not as steep as the normative model implies. People inferred that $P(c = 1) = .78$, $P(m = 1) = .73$, and $P(e = 1) = .67$, but given their belief that $P(c = 1) = .78$, the other two inferences should have been $P(m = 1) = .58$, and $P(e = 1) = .44$. In sum, people were insufficiently sensitive to the strengths.

Common Cause: $E_1 \leftarrow C \rightarrow E_2$

Inferring the Cause From Effects: $P(C|E_1, E_2)$

Assuming positive causal relationships, the more effects that are present, the more likely the cause is to be present. Rehder and Burnett (2005) confirmed that research participants demonstrate this effect. However, there are no results yet on how close this inference is to the normative calculations. In particular, consider a case in which C influences three effects, all with the same strength. The difference in the likelihood of C being present when only one versus two of the effects are present should be larger than the difference between two versus three of the effects. This pattern of reasoning is not apparent in Rehder and Burnett's (2005) experiments, although their experiments do not provide a strong test for this effect because participants did not know the precise parameters.

Inferring One Effect From Another Effect: $P(E_1|E_2)$

Two effects of a common cause should in general be correlated (Equation 4). The reason is simply that when the common cause C is present, assuming positive causal relations, then all the effects are more likely to be present, but when the common cause is absent, all the effects are more likely to be absent.

Equation 4

$$P(e_1 = 1|e_2 = 1) = \frac{P(e_1 = 1|c = 1)P(e_2 = 1|c = 1)P(c = 1) + P(e_1 = 1|c = 0)P(e_2 = 1|c = 0)P(c = 0)}{P(e_2 = 1|c = 1)P(c = 1) + P(e_2 = 1|c = 0)P(c = 0)}$$

Waldmann and Hagmayer (2005) performed a study in which participants were given a common cause structure and experienced a series of learning trials during which they observed all three variables. At the end participants inferred $P(e_1 = 1|e_2 = 1) > P(e_1 = 1|e_2 = 0)$, reflecting transitivity. However, this is not surprising because out of the 20 learning trials, in all but two, E_1 and E_2 had the same value. More impressive was that these inferences were sensitive to $P(c = 1)$ and the causal strengths. However, one problem with this study for our purposes was that participants experienced learning trials in which they observed C, E_1, and E_2. Thus, it is possible that when they were inferring $P(e_1 = 1|e_2 = 1)$, they were merely making a direct inference from E_2 to E_1 rather than reasoning from E_2 up to C and then back down to E_1.

Hagmayer and Waldmann (2000; see also Waldmann, Cheng, Hagmayer, & Blaisdell, 2008) conducted a similar study, but on each learning trial participants either observed C and E_1 or C and E_2. They had to infer the correlation between the two effects based on the causal model. At the end, participants inferred $P(e_1 = 1|e_2 = 1)$ and $P(e_1 = 1|e_2 = 0)$, and later these were converted to a *phi* correlation coefficient. This condition was also compared to a common effect structure, $C_1 \rightarrow E \leftarrow C_2$, in which participants learned about C_1 and E or C_2 and E and later judged the correlation between C_1 and C_2. Unlike a common cause, a common effect structure implies no correlation between C_1 and C_2.

In Experiment 1,[1] people's estimates of the correlation ($r = .29$) were much weaker than the true correlation ($r = .62$) and were not significantly different from the common effect control condition. In Experiment 2, the inferences also were quite low ($r = .26$) compared to the normative value ($r = .44$)[2] and, again, not different from a control condition. Perales et al. (2004) reported a similar set of experiments. Their participants did infer correlations between E_1 and E_2, and they gave higher correlations in the common cause than in the common effect condition. However, in

[1] Hagmayer and Waldmann (2000) also collected a separate "implicit" measure that was closer to the normative value. However, this measure again might not reflect pure reasoning from E_2 to C and then to E_1 because participants *observed* C and then predicted E_1 and E_2.

[2] This is a different value than the one cited in the original article because of a slight error in calculating ΔP.

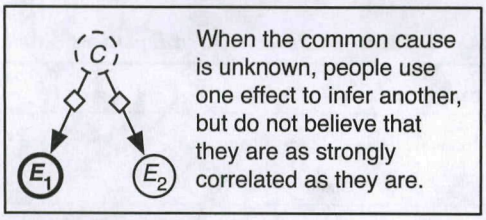


Figure 9. Summary of the $P(E_1|E_2)$ inference. See Figure 4 for notational conventions.

some of the conditions, particularly those with deterministic links, the inferred correlations were considerably lower than the normative calculation (although they used an unusual correlation rating scale).

One final experiment tested this inference in a different way. Von Sydow et al. (2010, Experiment 2; see also the discussion of transitivity in causal chains above) told participants about the common cause structure and had them observe 192 learning trials of all three variables. Recall that even though there were correlations between C and E_1 and C and E_2, there was zero correlation between E_1 and E_2 (i.e., the learning trials violated the Markov condition). In contrast to the chain structure in which people inferred transitivity, their judgments of $P(E_2|E_1)$ for the common cause structure implied no transitivity. In sum, these experiments suggest that people do not always believe that effects of a common cause are correlated, even though causal Bayesian networks imply that they usually are. See Figure 9 for a summary of the $P(E_1|E_2)$ inference.

Inferring E_1 After an Intervention on E_2

Waldmann and Hagmayer (2005; Experiments 3 and 4) also had participants infer E_1 after an intervention on E_2; $P(e_1 = 1|\text{set } e_2 = 1)$ or $P(e_1 = 1|\text{set } e_2 = 0)$. An intervention on E_2 severs the link from C to E_2, so the only way to infer E_1 is directly from C. Waldmann and Hagmayer (2005) found that when C had a higher base rate, $P(e_1 = 1|\text{set } e_2 = 1)$ was higher. This finding suggests that people have an understanding of what an intervention means in terms of causal structures and that they are able to perform inferences on the remaining causal structure. Manipulating the strength of C on E_1 also had some effect on the inference of E_1.

However, participants did not answer these questions entirely normatively. First, when the base rate of C and the strength of C on E_1 were manipulated, the inferences did not change as much as the normative model predicts they should change. Additionally, participants predicted that E_1 was more likely to be present when E_2 was intervened upon and set to 1 compared to 0, even though the intervention implies that E_2 is irrelevant for inferring E_1.

Hagmayer and Sloman (2009) tested whether people would recommend an action intervening on E_2 to produce a change in E_1. Surprisingly, there were some participants who recommend such an intervention. However, as with all studies using real-world knowledge, it is hard to know if these participants had additional beliefs, not explicit in the instructions, which would justify such an intervention (e.g., perhaps they believed that there might be an additional link $E_2 \rightarrow E_1$). See Figure 10 for a summary of the $P(E_1|\text{set } E_2)$ inference.

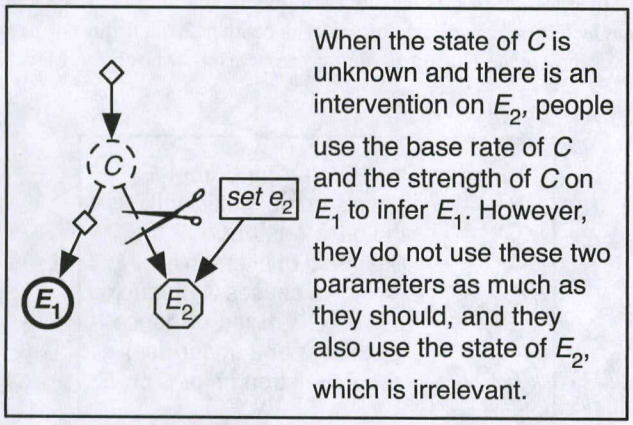


Figure 10. Summary of the $P(E_1|\text{set } E_2)$ inference. See Figure 4 for notational conventions.

One Link $C \rightarrow E$

In this section we discuss how people use the three parameters of a one-link causal structure, $P(c = 1)$, S_C, and S_A, when performing various inferences.

Inferring E Given C

As can be deduced from Table 1 row 1, $P(e = 1|c = 1) = 1 - (1 - S_C)(1 - S_A)$. Fernbach, Darlow, and Sloman (2011; Experiment 2) tested whether people are sensitive to S_A using scenarios involving generic real-world events. For example, the unpopularity of the mayor of a city (C) could cause the mayor's new policy to be unpopular (E), but a policy could be unpopular for other reasons even if the mayor is popular (A). Participants were asked three questions that defined the parameters of the one-link structure: the probability that a mayor of a major city is unpopular, $P(c = 1)$, the probability that the mayor's unpopularity would cause his or her new policy to be unpopular, S_C, and the probability that a new policy would be unpopular even if the mayor is popular, $P(e = 1|c = 0) = S_A$. Fernbach et al. then used these three parameters to predict how participants would judge $P(e = 1|c = 1)$, the probability of a policy's being unpopular given that the mayor is unpopular.

Fernbach et al.'s (2011) participants' inferences were mainly determined by the strength of the primary cause, S_C, and were not correlated with their beliefs about S_A. Their inferences of $P(e =$

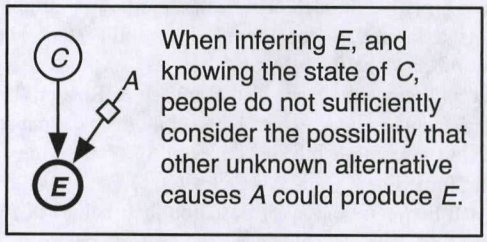


Figure 11. Summary of the $P(E|C)$ inference. See Figure 4 for notational conventions.

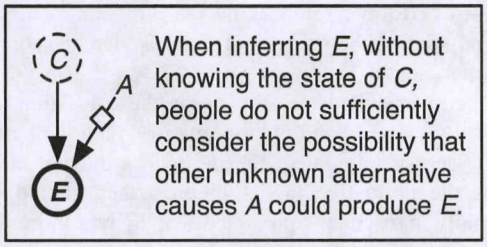


Figure 12. Summary of the $P(E)$ inference. See Figure 4 for notational conventions.

$1|c_1 = 1)$ were also 8% lower than the normative model (calculated using each participant's responses to the other three questions). Fernbach et al. attribute both of these results to participants' failure to consider the possibility that A could produce E. An alternative interpretation is that unless A is explicitly mentioned, people interpret the question $P(e = 1|c = 1)$ to be asking for S_C.

Of course, whenever real-world stimuli are used it is hard to know if participants have additional causal beliefs not picked up by the experimenters. Fernbach and Rehder (2012; Experiments 1 and 2) tested the same phenomenon, but with artificial stimuli. For example, they said that iodized helium (C) causes stars to be very hot (E). They also told participants the strength of the causal relationship, S_C and the likelihood that some other factor caused the same effect, $P(e = 1|c = 0) = S_A$. When they manipulated both factors in a 2 × 2 design, participants clearly made use of S_C but were not at all sensitive to S_A.

Fernbach and Rehder (2012) also asked participants to estimate $P(e = 1|c = 0)$. This should have been extremely easy because participants were explicitly told the likelihood of some other cause producing the effect; S_A. However, they were insensitive to variation in S_A. This is odd given that participants literally had this piece of information right in front of them—it was a parameter, not an inference. In sum, people appear to view S_A as less relevant than it is in reality for both $P(e = 1|c = 1)$ and $P(e = 1|c = 0)$. See Figure 11 for a summary of the $P(E|C)$ inference.

Inferring the Effect

As shown in Table 1 row 1, $P(e = 1) = 1 - (1 - S_C P(c = 1))(1 - S_A)$. Fernbach et al. (2011; see description above) collected judgments of $P(e = 1)$, and we analyzed the results by calculating what $P(e = 1)$ should have been given their participants' average estimates of $P(c = 1)$, S_C, and S_A. Just as for the $P(e = 1|c = 1)$ inference, their participants' inferences of $P(e = 1)$ were 9% lower than the normative model. This underprediction might be explained as a failure to consider the possibility that alternative causes could produce the effect.

Rehder and Kim (2010; see also Fernbach & Rehder, 2012) also asked participants to infer $P(e = 1)$. Overall, their participants were sensitive, but not sufficiently sensitive, to S_A. For example, in one condition (Experiment 2 in Appendix C) $S_C = .75$ and S_A was manipulated between 0 and .75. Based on their beliefs of $P(c = 1)$, participants' inferences should have changed from .56 to .88, but they changed only from .69 to .79. See Figure 12 for a summary of the $P(E|C)$ inference.

Table 3
Direction of Influence of Parameters in Equations 5 and 6

Parameter	$P(c = 1\|e = 1)$	$P(c = 1\|e = 0)$
$P(c_1 = 1)$	↑	↑
S_C	↑	↓
S_A	↓	—

Note. Parameter has an increasing (↑), decreasing (↓), or no effect (—) on the inference.

Inferring C Given E

Inferring a cause given knowledge of an effect is called a "diagnostic inference" as an analogy to medical diagnosis in which a disease (cause) is sought to explain a set of symptoms (effect). Equations 5 and 6 show these inferences, and the Table 3 shows the directions of the influences of the parameters assuming positive strengths. (See Meder, Mayrhofer, and Waldmann, 2009, for a modified normative framework for inferring $P(c = 1|e = 1)$ when the causal structure is not known a priori.) In the following three sections, we separately evaluate the evidence of whether people are sensitive to the three parameters for inferring $P(c = 1|e = 1)$.

Equation 5

$$P(c = 1|e = 1) = \frac{S_C + S_A - S_C S_A}{S_C + S_A/P(c = 1) - S_C S_A}$$

Equation 6

$$P(c = 1|e = 0) = \frac{1 - S_C}{1/P(c = 1) - S_C}$$

Use of $P(c = 1)$. When inferring causes from effects, people notoriously exhibit base rate "neglect" or "underappreciation"; in other words, they fail to use $P(c = 1)$ to the extent dictated by Bayes' rule (e.g., Bar-Hillel, 1980; Eddy, 1982; Kahneman & Tversky, 1972; Koehler, 1996). In contrast, others have suggested that when people learn the parameters from experience instead of being told the parameters, their inferences are closer to the correct Bayesian calculation (e.g., Christensen-Szalanski & Beach, 1982; Gigerenzer & Hoffrage, 1995; see also the discussion above on reasoning based on observed frequencies).

Irrespective of this debate, the previous research on "base rate neglect" often involved *statistical* dependencies and did not necessarily engage *causal* reasoning habits. Thus, we rely on Meder,

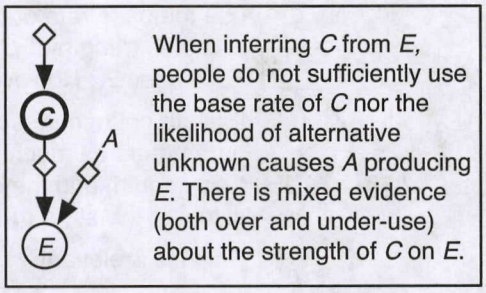


Figure 13. Summary of the $P(C|E)$ inference. See Figure 4 for notational conventions.

Hagmayer, and Waldmann's (2009) Experiment 2, which involved explicitly causal scenarios. Meder et al. (2009) taught participants a structure with four nodes and showed them a series of learning trials allowing them to learn the parameters from experience. Afterward, participants estimated $P(c = 1|e = 1)$. Even though the $C \rightarrow E$ link was part of a larger causal structure, the rest of the structure is irrelevant for this particular inference.

In one condition for which $P(c = 1) = .30$, $S_C = .35$, and $S_A = .57$, $P(c = 1|e = 1)$ should be .35. Participants' inferences were right on target. However, in another condition in which $P(c = 1) = .65$, $S_C = .80$, and $S_A = .24$, $P(c = 1|e = 1)$ should be .87; yet participants' inferences were only .51. Unfortunately for our purposes, all three parameters changed across the two conditions prohibiting a clear analysis of $P(c = 1)$. However, in the second condition $P(c = 1|e = 1)$ was lower than $P(c = 1)$, which should never happen with positive causal relations. This result reflects, at minimum, a misuse of the base rate.

Use of S_C. Meder et al. (2008) and Meder, Hagmayer, and Waldmann (2009) also examined the use of S_C: if $S_C > 0$, then $P(c = 1|e = 1) > P(c = 1|e = 0)$. One trend across all their experiments is that participants' inferences were not nearly as extreme as predicted. For example, in one experiment (Meder et al., 2008, Experiment 1), the normative probabilities were $P(c = 1|e = 1) = .95$ and $P(c = 1|e = 0) = .10$. However, participants' responses, converted to probabilities, were $P(c = 1|e = 1) = .76$ and $P(c = 1|e = 0) = .43$. Thus, even though the direction of the effect was correct, the estimates were "conservative."

Fernbach and Rehder (2012; Experiment 1) told their participants the parameters S_C and S_A and collected judgments of $P(c = 1|e = 1)$. Both S_C and S_A were manipulated in a 2×2 to be either strong or weak. Unfortunately, the third relevant parameter, $P(c = 1)$ was not provided to participants. We analyze this study two ways. First we assumed a plausible value of $P(c = 1) = .67$. Comparing the conditions when S_C was increased but S_A was held constant, we would only expect an increase of about .03–.06 in $P(c = 1|e = 1)$. However, participants inferences increased by about .20. Another way to analyze these data is to reverse-derive $P(c = 1)$ using the normative model given the supplied values of S_C, S_A, and the participants' average judgment of $P(c = 1|e = 1)$ for one condition. Then the normative inference for $P(c = 1|e = 1)$ can be derived for the other condition. This analysis also shows that participants' inferences varied *too much* based on the change in S_C.

Fernbach and Rehder's (2012) study also asked for inferences of $P(c = 1|e = 0)$. The increase in S_C led to a decrease in estimates of $P(c = 1|e = 0)$; this general pattern is normative. However, both methods of analysis reveal that the inferences of $P(c = 1|e = 0)$ changed *too little* based on the change in S_C. In sum, there does not appear to be a clear pattern of how S_C is utilized when inferring the state of a cause from an effect: There are complicated patterns of over- and underutilization of S_C relative to the normative model.

Use of S_A. Fernbach and Rehder (2012) told participants the parameters of the $C \rightarrow E$ structure, and they manipulated S_A. In their Experiment 1, the manipulation of S_A should have produced differences in $P(c = 1|e = 1)$ of about .12–.15, but participants inferred differences of only about .06. Although this was a significant difference, it is about half as much as is expected by the normative model.

Two studies have examined the impact of making the alternative cause A explicit when inferring the cause. The idea is to provide a reason for the times when the effect occurs without the observed cause. The standard task so far involves a $C \rightarrow E$ structure with an implicit alternative cause; we have interpreted $P(e = 1|c = 0)$ as S_A. Two studies have reframed the scenario as a common effect structure $C \rightarrow E \leftarrow A$, in which A is explicitly mentioned and the parameters of A are provided to participants.

Krynski and Tenenbaum (2007; Experiment 2) used the standard mammography base rate neglect problem (cancer causes a positive mammogram, but there can also be false positives). Participants were told the parameters $P(c = 1)$, S_C, and S_A, and they inferred $P(c = 1|e = 1)$. In one condition the false positive rate S_A was not explained; the possible causes were implicit. In another condition the wording explicitly mentioned a second cause of a positive mammogram result, a benign cyst.

The inferences were more normative in the condition in which the alternative cause was explicitly mentioned. About 42% of the inferences in the "explicit" benign cyst condition were right on target. In contrast, only 16% of the inferences in the "implicit alternative cause" condition were right on target. Krynski and Tenenbaum (2007) interpreted this result as showing that people have an easier time reasoning about explicit causes than about fundamentally stochastic causes (the unexplained false positive rate). However, this explanation is not entirely satisfying because it is unclear why people wouldn't just infer an additional cause for any one-link causal structure in which the effect occurred (positive mammogram) without the observed cause (malignant tumor).

Fernbach and Rehder (2012; Experiment 2) performed a similar manipulation making the alternative cause either implicit or explicit. Manipulating S_A had no effect in the explicit condition, and the effect in the implicit condition was much smaller than expected. In sum, multiple studies have found insufficient use of S_A. See Figure 13 for a summary of the $P(E)$ inference.

Common Effect: $C_1 \rightarrow E \leftarrow C_2$

In this section we focus on common effect structures when there are only two causes, C_1 and C_2, both of which are explicit in the model. This means that if both causes are absent then E must be absent because there is no alternative background cause A. In this case, still assuming a noisy-OR gate with no interactions, then $P(e = 1) = 1 - [1 - P(c_1 = 1)S_{C1}][1 - P(c_2 = 1)S_{C2}]$.

Discounting: $P(C_1|E)$ Versus $P(C_1|E,C_2)$

Here we continue the discussion on $P(c = 1|e = 1)$ that began in the section on $C \rightarrow E$ structures but now discuss this inference in relation to $P(c_1 = 1|e = 1, c_2 = 1)$. Assuming generative causal relationships (which we assume for this entire section), $P(c_1 = 1|e = 1) > P(c_1 = 1|e = 1, c_2 = 1)$. This inference is atypical: In most other structures the presence of one node increases the probability of another node (again assuming generative causal links). Alternatively, sometimes due to the Markov condition ("screening off"), one node is irrelevant to the probability of another node. But for a common effect structure, the presence of C_2 actually *decreases* the likelihood of C_1. This atypical reasoning pattern has been viewed as a key aspect of causal reasoning (see Khemlani & Oppenheimer, 2010, for a review).

We explain this pattern of reasoning using the Farming Scenario. In this scenario, an early frost and a tomato fruit-worm infestation are both sufficient to cause a poor tomato harvest;

Table 4
Example Data for Discounting

Row	Early frost (F)	Tomato fruit-worm infestation (W)	Poor tomato harvest (T)	Number of farms
A	1	1	1	10
B	1	0	1	90
C	0	1	1	90
D	0	0	0	810

$F \rightarrow T \leftarrow W$. Table 4 shows a hypothetical sample of 1,000 farms for which 10% experience an early frost and 10% experience a tomato fruitworm infestation.

Within the 190 farms that had a poor tomato harvest (rows A–C), 100 of them had a tomato fruit-worm infestation; $P(w = 1|t = 1) = .53$. But if we know that a farm had a poor harvest *and* that it also had an early frost (rows A and B), only 10 out of the 100 had a tomato fruit-worm infestation; $P(w = 1|t = 1, f = 1) = .10$. This phenomenon has been known in artificial intelligence (e.g., Pearl, 1988) as "explaining away," and in psychology as "discounting": knowing that the farm had an early frost explains away or discounts the possibility that the farm had an infestation.

More generally, the pattern of discounting can be conceived in the following way. Observing that E is present increases the probability that C_1 is present compared to its base rate; $P(c_1 = 1) < P(c_1 = 1|e = 1)$. Subsequently observing that C_2 is also present decreases the likelihood of C_1; $P(c_1 = 1|e = 1) > P(c_1 = 1|e = 1, c_2 = 1)$. If C_1 and C_2 are both sufficient to produce E, then the probability of C_1 falls all the way back down to its base rate; $P(c_1 = 1) = P(c_1 = 1|e = 1, c_2 = 1)$. However, if C_2 is weak and is unlikely to explain the presence of E, then the probability of C_1 still remains higher than its base rate; $P(c_1 = 1) < P(c_1 = 1|e = 1, c_2 = 1)$. See Equations 7 and 8 for the normative calculations, and see Table 5 for the direction of the influence of the variables in Equations 7 and 8. We now discuss empirical results related to discounting.

Equation 7

$$P(c_1 = 1|e = 1) = \frac{\frac{S_{C_1}}{P(c_2 = 1)} + S_{C_2} - S_{C_1}S_{C_2}}{\frac{S_{C_1}}{P(c_2 = 1)} + \frac{S_{C_2}}{P(c_1 = 1)} - S_{C_1}S_{C_2}}$$

Equation 8

$$P(c_1 = 1|e = 1, c_2 = 1) = \frac{S_{C_1} + S_{C_2} - S_{C_1}S_{C_2}}{S_{C_1} + \frac{S_{C_2}}{P(c_1 = 1)} - S_{C_1}S_{C_2}}$$

The prototypical "discounting" effect: $P(c_1 = 1|e = 1, c_2 = 1) < P(c_1 = 1|e = 1)$. Morris and Larrick (1995) defined discounting as the relationship between $P(c_1 = 1|e = 1)$ and $P(c_1 = 1|e = 1, c_2 = 1)$ and asked whether people discount normatively. There is a rich history of research on discounting within social psychology. However, most of these studies did not present people with all the parameters of the model, nor did they assess the parameters that participants were intuitively using, so that a normative analysis is not possible. To answer this question, Morris and Larrick (1995; pp. 340–341) conducted a study using a classic discounting scenario in which participants were told that they would read essays written by other students about Castro's regime in Cuba; half of the writers were randomly assigned to write essays that were pro- or anti-Castro (E. E. Jones & Harris, 1967). In terms of the causal structure framework, one of the potential causes, C_1, was whether the writer's *personal attitude* was pro- or anti-Castro. The second potential cause, C_2, was whether the writer was *assigned* to write an essay that was pro- or anti-Castro. The effect, E, was whether the essay was pro- or anti-Castro.

Participants first judged the following four parameters: $P(c_1 = 1)$, the prior probability of the writer having a pro-Castro attitude, $P(c_2 = 1)$, the prior probability of a writer being assigned to write a pro-Castro essay, $P(e = 1|c_1 = 1, c_2 = 0) = S_{C1}$, the probability that a person with a pro-Castro attitude would write a pro-Castro essay even if he or she was assigned to write an anti-Castro essay, and $P(e = 1|c_1 = 0, c_2 = 1) = S_{C2}$, the probability that a person with an anti-Castro attitude would write a pro-Castro essay if he or she was assigned to write an pro-Castro essay. After reading the essay, which was always pro-Castro, the participants rated the probability that the writer had a pro-Castro attitude $P(c_1 = 1|e = 1)$. Finally, participants were told that the writer was *assigned* to write a pro-Castro essay, and the participants judged again whether the writer's attitude was pro-Castro, $P(c_1 = 1|e = 1, c_2 = 1)$.

Morris and Larrick (1995) found the normative discounting effect, $P(c_1 = 1|e = 1) = .35 > P(c_1 = 1|e = 1, c_2 = 1) = .30$. In fact, the inference of $P(c_1 = 1|e = 1)$ was close to the normative calculation of .36 based on participants' own beliefs about the parameters. However, the $P(c_1 = 1|e = 1, c_2 = 1) = .30$ inference was numerically higher than the normative calculations (.26), though not significantly so. Thus, it seems that participants did discount, though only about half as much as they should have.

Fernbach and Rehder (2012; Experiment 3, "present" condition) told participants a hypothetical scenario about a common effect structure, instructed them about S_{C1} and S_{C2}, and asked them to infer $P(c_1 = 1|e = 1)$ and $P(c_1 = 1|e = 1, c_2 = 1)$. In one condition in which S_{C2} was strong, there is a slight trend for $P(c_1 = 1|e = 1, c_2 = 1) < P(c_1 = 1|e = 1)$. Yet in another condition in which S_{C2} was weak, there was a slight trend in the opposite direction. Discounting should normatively be greater when S_{C2} is stronger, but it should never go in the opposite direction so long as the two causes are independent (see the next section). Unfortunately because participants were not told specific values for $P(c_1 = 1)$ and $P(c_2 = 1)$ we cannot quantitatively compare these inferences to the normative model.

Rehder (2012) told participants about a common effect structure without the parameters and then had them choose which one is

Table 5
Direction of Influence of Parameters in Equations 7 and 8

Parameter	$P(c_1 = 1\|e = 1)$	$P(c_1 = 1\|e = 1, c_2 = 1)$
$P(c_1 = 1)$	↑	↑
S_{C1}	↑	↑
$P(c_2 = 1)$	↓	—
S_{C2}	↓	↓

Note. Parameter has an increasing (↑), decreasing (↓), or no effect (—) on the inference.

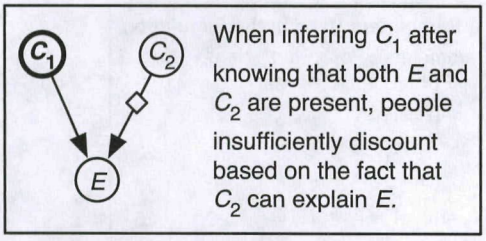


Figure 14. Summary of discounting. See Figure 4 for notational conventions.

higher (or equal): $P(c_1 = 1|e = 1,c_2 = 1)$ versus $P(c_1 = 1|e = 1)$. Across two experiments participants were either more likely to choose the former—the opposite of discounting—or there was not a significant difference.

In sum, relatively few studies that have examined discounting allow for comparisons to the normative model. Out of those that do, discounting appears to be weak, and sometimes the inferences go in the opposite direction of discounting.

A related discounting effect: $P(c_1 = 1|e = 1,c_2 = 1) < P(c_1 = 1|e = 1,c_2 = 0)$. Several studies have compared inferences of C_1 when C_2 is present versus absent. Hagmayer and Waldmann (2007) and Luhmann and Ahn (2007; Experiment 3) presented participants with a series of learning trials; on each trial they observed C_2 and E and judged whether C_1 was present or absent. Fernbach and Rehder (2012; absent versus unknown conditions in Experiment 3) and Rehder (2011; independent condition) told people the parameters S_{C1} and S_{C2} and had them make judgments of $P(c_1 = 1|e = 1, c_2 = 1)$ and $P(c_1 = 1|\, e = 1,c_2 = 0)$.

For all these studies, participants' inferences did exhibit the expected asymmetry $P(c_1 = 1|\, e = 1,c_2 = 1) < P(c_1 = 1|\, e = 1,c_2 = 0)$. Unfortunately, in all these studies $P(c_1 = 1)$ was not identified, so quantitative comparisons to the normative model for $P(c_1 = 1|\, e = 1,c_2 = 1)$ were not possible. There was, however, an unexpected pattern. Because there are only two possible causes of E, $P(c_1 = 1|\, e = 1,c_2 = 0)$ should equal 1. In Hagmayer and Waldmann's (2007) study (Experiment 1), participants' inferences of $P(c_1 = 1|\, e = 1,c_2 = 0)$ were close to 1, but in Luhmann and Ahn's (2007) study and Fernbach and Rehder's (2012) study they were around 0.75. A possible interpretation would be that participants inferred that there was another unobserved cause of E. Inferring another unobserved cause could also dampen the standard discounting effect $P(c_1 = 1|e = 1,c_2 = 1) < P(c_1 = 1|e = 1)$. However, if participants in these studies had inferred unobserved generative causes, they should also have given higher ratings of $P(e = 1|c_1 = 1)$ than would be expected from the two known causes. At least in Fernbach and Rehder's (2012) study, this was not the case.

Sussman and Oppenheimer's (2011) investigation involved three variables representing plumbing parts (e.g., tightness of a clamp, amount of water flowing through a spout); they also tested discounting. The authors found no discounting in Experiment 1, and in Experiment 2, discounting was less than predicted by the normative model. See Figure 14 for a summary of discounting.

The influence of S_{C2} on $P(c_1 = 1|e = 1)$ and $P(c_1 = 1|e = 1,c_2 = 1)$. Both $P(c_1 = 1|e = 1)$ and $P(c_1 = 1|e = 1,c_2 = 1)$ should decrease with higher values of S_{C2}; the stronger that C_2 is, the more sufficient that C_2 is to explain the presence of E and thus the less that C_1 is needed to explain E. Fernbach and Rehder's (2012) participants were told about the common effect structure and told the parameters S_{C1} and S_{C2}. Participants' inferences of $P(c_1 = 1|e = 1,c_2 = 1)$ were lower when S_{C2} was higher (Experiment 3; Present condition). However participants' inferences of $P(c_1 = 1|e = 1)$ were not sensitive to S_{C2} (Experiment 2, explicit condition and Experiment 3, unknown condition). But, these inferences cannot be quantitatively compared to the normative model because two parameters, $P(c_1 = 1)$ and $P(c_2 = 1)$, were not known.

Discounting when two causes are correlated. In the previous discussion of discounting on a common effect model, $C_1 \rightarrow E \leftarrow C_2$, the two causes were assumed to be independent, in which case $P(c_1 = 1) \leq P(c_1 = 1|e = 1) \geq P(c_1 = 1|e = 1,c_2 = 1)$. Here we consider instances when the two causes are correlated (see Figure 15B), in which case these inequalities do not necessarily hold. Instead of presenting equations, we explain discounting with correlated causes using Figure 15. C_1 and C_2 could be correlated if there is an underlying common cause or a direct link between C_1 and C_2; the inferences in Figure 15A are derived assuming an additional link $C_1 \rightarrow C_2$ with the joint probability $P(C_1,E,C_2) = P(E|C_1,C_2)P(C_2|C_1)P(C_1)$.

One easy way to think about discounting is to consider how the inference about C_1 changes after first learning that $e = 1$, and again after also learning that $c_2 = 1$. Thus, one should read Figure 15A from left to right. The "independent" line in Figure 15A shows a typical discounting pattern when the two causes are independent. Learning that E is present increases the probability of C_1. Then, learning that C_2 is also present decreases the probability of C_1.

The "positive" line in Figure 15A shows how the pattern of inferences involved in discounting is affected when the two causes are positively correlated. First, $P(c_1 = 1|e = 1)$ is higher when they are positively correlated compared to when they are independent. To understand why, consider the common effect with correlated causes structure in Figure 15B. Learning that $e = 1$ increases the probability of C_1 through the $C_1 \rightarrow E$ link, and it also indirectly increases the probability of C_1 through the $C_1 — C_2 \rightarrow E$ route. Subsequently learning that $C_2 = 1$ results in a smaller drop in the probability of C_1 compared to the structure with independent causes. The reason is that learning that $c_2 = 1$ decreases the probability of C_1 through normal discounting (the "bottom" path $C_1 \rightarrow E \leftarrow C_2$) but *increases* the probability of C_1 through the $C_1 — C_2$ route.

Now consider how discounting is influenced by a negative correlation between the two causes. Learning that $e = 1$ increases the probability of C_1 through the direct link $C_1 \rightarrow E$ but decreases the probability of C_1 through the path $C_1 — C_2 \rightarrow E$. This means that $P(c_1 = 1|e = 1)$ is lower compared to when the causes are independent (line Negative 1 in Figure 15). In fact, if the $C_1 — C_2 \rightarrow E$ path is strong and $C_1 \rightarrow E$ is weak (Negative 2 line), then it is possible for $P(c_1 = 1) > P(c_1 = 1|e = 1)$.

Subsequently learning that $c_2 = 1$ results in a greater drop from $P(c_1 = 1|e = 1)$ to $P(c_1 = 1|e = 1,c_2 = 1)$, compared to when the causes are independent (compare the Independent vs. Negative 1 lines because they have similar parameters). Learning that $c_2 = 1$ decreases the probability of C_1 through normal discounting (the "bottom" path $C_1 \rightarrow E \leftarrow C_2$) and also directly decreases the probability of C_1 through the $C_1 — C_2$ path.

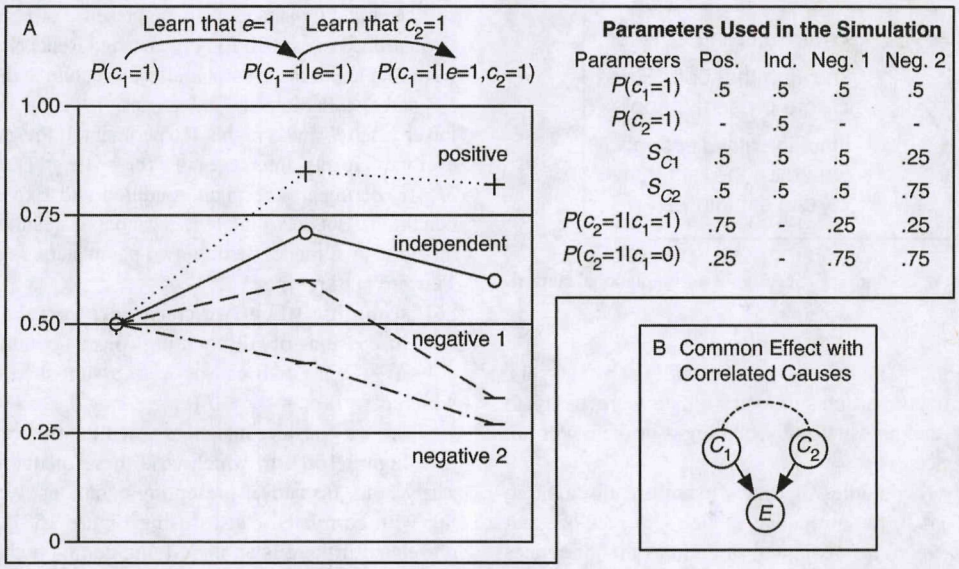


Figure 15. Discounting when causes are dependent versus independent; the graph plots $P(c_1 = 1)$ as the states of E and C_2 are learned. Pos. = positive; Neg. = negative; Ind. = independent.

Morris and Larrick (1995; Experiment 2) tested whether people use the correlation between C_1 and C_2 in a discounting task. Again, the participants read essays about Castro's regime, and they inferred whether the writer was pro-Castro (C_1) given that the essay was pro-Castro ($e_1 = 1$), both before and after learning that the writer was assigned to write a pro-Castro essay ($c_2 = 1$). In the independent condition, writers were supposedly assigned to write pro- or anti-Castro essays randomly. In the positive versus negative correlation conditions, pro-Castro writers were likely to be assigned to write pro-Castro essays (positive condition) or anti-Castro essays (negative condition). Consistent with the normative standard, participants discounted most strongly [$P(c_1 = 1|e = 1)$ vs. $P(c_1 = 1|e = 1, c_2 = 1)$] in the negative correlation condition, least strongly in the positive correlation condition, and at an intermediate amount in the independent condition.

However, one aspect of the results, not discussed by Morris and Larrick (1995), was that across all conditions and for both judgments of $P(c_1 = 1|e = 1)$ and $P(c_1 = 1|e = 1, c_2 = 1)$, participants tended to provide lower estimates compared to the normative standard. This underprediction resulted in some surprising patterns of reasoning. In the negative correlation condition, participants' average inference of $P(c_1 = 1|e = 1) = .38$ was lower than their inference of $P(c_1 = 1) = .48$ (we do not know if it was significantly lower); $P(c_1 = 1|e = 1)$ should have been .51. In the independent condition, the two inferences were essentially equal, $P(c_1 = 1|e = 1) = P(c_1 = 1) = .48$, even though $P(c_1 = 1|e = 1)$ should have been .63. In the positive correlation condition $P(c_1 = 1|e = 1) > P(c_1 = 1)$, although the difference was not as large as expected. In sum, this experiment suggests that people are remarkably normative in their overall pattern of discounting, but the inferences were biased to be low. These low inferences might be explained by participants underweighting their own prior on C_1 or by their own strength of C_1. See Figure 16 for a summary of the $P(E|C_1, C_2)$ inference.

Summary of discounting. A number of studies have demonstrated that people sometimes discount the likelihood of one cause when another cause is known to have occurred and is sufficient to explain the presence of the effect. People are even sensitive to the correlation between the two causes. However, there are also a number of findings in which discounting was considerably smaller than the amount implied by the normative model, in which there was no discounting at all, or in which the inferences went in the opposite direction of discounting. Clearly there are many remaining empirical questions about discounting.

Use of the Base Rates in Diagnostic Judgments

Reips and Waldman (2008) conducted a study of diagnostic learning when there were two diseases (causes) that both caused the same symptom (effect). Both diseases always caused the symptom, so the diagnostic judgment $P(c_1 = 1|e = 1)$ should perfectly reflect the frequency of the diseases. Participants learned from experience that C_1 was three times more common than C_2, and they could accurately report the base rates. Although their judgments of $P(c_1 = 1|e = 1)$ were greater than $P(c_2 = 1|e = 1)$, the difference was not close to the expected 3:1 ratio. Similar to the results for the $C \rightarrow E$ structure, this result implies undersensitivity to the base rates.

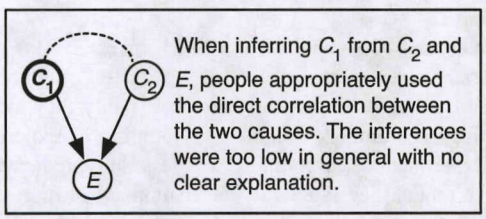


Figure 16. Summary of discounting with correlated causes. See Figure 4 for notational conventions.

$P(c_1 = 1|e = 0, c_2 = 1)$ Versus $P(c_1 = 1|e = 0, c_2 = 0)$ With Conjunctive Causes

So far we have only discussed scenarios in which C_1 and C_2 combine through a Noisy-OR rule. Rehder (2011) investigated how people reason about probabilistic *conjunctive* causes; when C_1 and C_2 are both present, the combination can cause E to be present, and there is also another independent unobserved cause of E. In this case, $P(c_1 = 1|e = 0,c_2 = 1,) < P(c_1 = 1|e = 0,c_2 = 0)$. When the effect is absent, if $c_2 = 1$, then C_1 is probably absent (if it was present then E would probably have been present). However, if $c_2 = 0$, then c_1 could be either 0 or 1.

Rehder (2011) presented participants with a common effect structure and a *conjunctive causes* cover story, then told participants the causal strength parameters, and asked them to infer $P(c_1 = 1|e = 0, c_2 = 1)$ and $P(c_1 = 1|e = 0, c_2 = 0)$. He found the predicted asymmetry. Additionally, he found that the inferences of $P(c_1 = 1|e = 0,c_2 = 0)$ were low, despite the fact that C_1 was described as usually being present. This pattern probably reflects misuse of the base rate of C_1.

Inferring E From Multiple Causes: $P(E|C_1,C_2)$

Fernbach and Rehder (2012; Experiment 3, "present" condition) conducted a study in which participants were told about a common effect structure, were told about the strength parameters [$S_{C1} = 0.6$ and $S_{C2} = .25$ versus 0.75], and were asked to infer $P(e = 1|c_1 = 1, c_2 = 1)$. Although this inference was higher when S_{C2} was higher, the difference was not as large as it should have been. The normative model predicts .9 versus .7, a difference of .2, but their participants inferred only a difference of about .07. People do not use the strength parameters as strongly as they should. See Figure 17 for a summary of the $P(E|C_1, C_2)$ inference.

Counterfactual Questions: $P(c_1 = 1|$ If e Had Been 0 Instead of 1)

So far we have discussed inferences based on observations and interventions. Here we discuss a third type of inference, counterfactuals. Counterfactuals involve first observing the states of the nodes and then asking a question about what would have been true if the actual conditions had not all occurred. For example, suppose one year on the farm (see Figure 1) there was neither an early frost nor an infestation, and there was a good tomato harvest. A counterfactual could be "What is the likelihood of a good tomato harvest if there had been an early frost?"

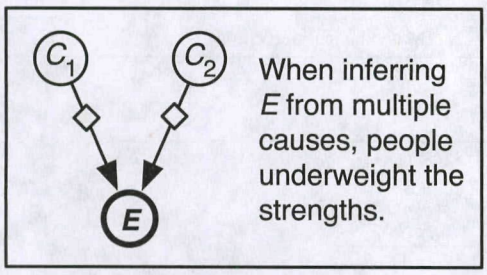


Figure 17. Summary of the $P(E|C_1, C_2)$ inference. See Figure 4 for notational conventions.

One possible solution is to treat counterfactuals as observations; $P(good\ harvest \mid early\ frost)$. The problem with this interpretation is that it discards our knowledge that the farm did not have an infestation. Pearl (2000) proposed that in many situations counterfactuals could be interpreted as interventions. In this case, the counterfactual would be interpreted as $P(good\ harvest \mid early\ frost, no\ infestation)$.

Consider a different counterfactual: "What is the likelihood of an early frost if there had been a poor tomato harvest?" (Remember that we know that there was neither an early frost nor an infestation, and there was a good harvest.) According to the intervention account, the intervention is on the poor harvest, which would mean that we maintain the belief that there was neither an early frost nor an infestation. The potential problem with this account is that one might think the following: "if there had been a poor harvest, there is a decent chance that it is due to an early frost." The intervention account does not allow for reasoning "upstream."

Hiddleston (2005) proposed another, more complicated, way to represent counterfactuals that involves thinking about the possible minimal changes to the causal network in which the counterfactual is true, but all the other nodes are "minimally" different from their actual states. In contrast to the intervention account, a minimal change could involve changes in nodes "upstream" of the counterfactual variable.

Rips (2010; see also Sloman & Lagnado, 2005) tested how people interpret counterfactuals on a common effect structure. Across a variety of conditions, Rips found that none of the strategies (observations, interventions, or "minimal-networks") by itself could account for all the results; he eventually proposed a modified version of the minimal-networks approach. In sum, it is not yet clear exactly how people interpret counterfactuals, and there is still expert disagreement on the normative interpretation of counterfactuals.

Conditional, "If . . . Then" Reasoning and Acceptability of Logical Arguments

There is a large literature on people's inferences involving propositions stated in an "If . . . Then" syntactic format (Evans & Over, 2004, provides an excellent introduction). Many such sentences refer to causal relationships, and some philosophers and experimentalists have proposed that conditional statements "If p, then q," are often interpreted probabilistically as $P(q = 1|p = 1)$ (Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003; Over, Hadjichristidis, Evans, Handley, & Sloman, 2007; see Bennett, 2003, on "The Ramsey Test").

Furthermore, the logical rules of inference (Modus Ponens, Modus Tollens, Denying the Antecedent, and Affirming the Consequent) can also be interpreted as probabilistic inferences instead of logical. For example, consider the premise "If $c = 1$, then $e = 1$" (i.e., $C \rightarrow E$). "Affirming the consequent" is the inference "$e = 1$, therefore $c = 1$." Logically this inference is invalid; however, consider how this inference might be viewed from causal structure perspective (Fernbach & Erb, in press; see Liu, Lo, & Wu, 1996; Oaksford, Chater, & Larkin, 2000, for other probabilistic accounts). First, in instances when the premise "If $c = 1$, then $e = 1$" refers to a causal relationship, $C \rightarrow E$, one may extend the structure with background knowledge and include $P(c = 1)$ and S_C

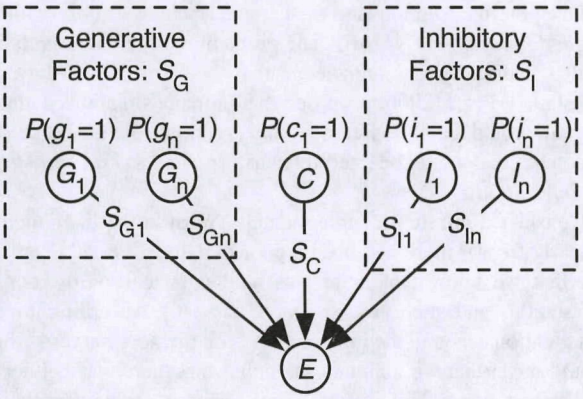


Figure 18. Causal structure for conditional "If . . . then" reasoning.

as well as other generative or inhibitory causes of E to form a structure like that in Figure 18. Second, assessing the acceptability of the inference "$e = 1$, therefore $c = 1$" could be interpreted as a request for the inference $P(c = 1|e = 1)$. Thereby, the four canonical forms of logical argumentation can be reframed as conditional probability inferences. Table 6 maps between the logical and probabilistic interpretations of these inferences and gives mathematical derivations of the probabilistic inferences on the structure in Figure 18. For comprehensibility, we talk about these logical inferences as interchangeable with conditional probability notation, even though many of the experiments actually had people judge the validity or acceptability of the logical arguments.

Although not explicitly framed in terms of Causal Bayesian Networks, a number of studies have examined the effect of manipulating various parameters of the causal structure, the number or strength of alternative generative or inhibitory causes, on the perceived validity of logical arguments. We use S_G and S_I to denote the total likelihood of alternative causes generating versus inhibiting E. Higher S_G reflects lower necessity and higher S_I reflects lower sufficiency of the $C \rightarrow E$ relation. There are *four* classic and robust findings (see superscript notation [a] in Table 6; e.g., Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys, Schaeken, & d'Ydewalle, 2003a; Quinn & Markovits, 1998). Increasing S_I leads to (a) lower judgments of Modus Ponens and (b) lower judgments of Modus Tollens. For example, given the conditional from Cummins (1995), "If John studied hard, then he did well on the test," it is possible to think of many possible disabling conditions (e.g., the test was very hard), which leads to a lower endorsement of Modus Ponens "John studied hard . . . therefore he did well on the test." Additionally, increasing S_G leads to (c) lower judgments of Denying the Antecedent and (d) lower judgments of Affirming the Consequent. These classic findings make perfect sense in terms of inferences on a causal network, and Fernbach and Erb (in press) have proposed a causal network framework to model such effects (although they used a slightly different structure than Figure 18). (Note that these classic studies predicted no effect of S_G on Modus Ponens and Modus Tollens and no effect of S_I on Denying the Antecedent and Affirming the Consequent.)

This causal model account of conditional inference has a number of benefits. First, it clarifies the features of the causal scenario that matter: the number, base rates, and strengths of the alternative generative and inhibitory causes (Fernbach & Erb, in press). We group these factors together as S_G and S_I. In addition, one's belief in the integration function would be critical, although here we are only discussing noisy-OR.

Second, this account makes many of the same predictions as those made in the classic studies (e.g., Cummins, 1995, see the superscript notation [a] in Table 6). In fact, it explains why S_I is predicted to have no effect on Affirming the Consequent; it falls out of the equation in row 4. However, we note that some studies have found a positive effect S_I (Beller, 2006; De Neys, Schaeken, & d'Ydewalle, 2002, Experiment 2, De Neys, Schaeken, & d'Ydewalle, 2003b).

Third, this account makes three different predictions than the standard ones (see the arrows in Table 6 not marked with a superscript). First, increasing S_G should increase the acceptance of MP, $P(e = 1|c = 1)$. People sometimes ignore implicit alternative generative causes for $P(e = 1|c = 1)$ judgments (see the $C \rightarrow E$ section), and Fernbach and Erb (in press) did not include them in their model, although two studies found this

Table 6
Logical and Probabilistic Interpretations of the Acceptability of Arguments

Logic			Probability			
					Influence of increasing	
Logical name	Inference	Valid	Inference	Mathematical derivation	S_G	S_I
MP: Modus ponens	$c = 1 \therefore e = 1$	Yes	$P(e = 1\|c = 1)$	$(S_C + S_G - S_C S_G)(1 - S_I)$	↑	↓[a]
MT: Modus tollens	$e = 0 \therefore c = 0$	Yes	$P(c = 0\|e = 0)$	$1 - \frac{1-(S_C+S_G-S_C S_G)(1-S_I)}{1/P(c=1)-(S_C+S_G/P(c=1)-S_C S_G)(1-S_I)}$	↓	↓[a]
DA: Deny antecedent	$c = 0 \therefore e = 0$	No	$P(e = 0\|c = 0)$	$1 - S_G(1 - S_I)$	↓[a]	↑
AC: Affirm consequent	$e = 1 \therefore c = 1$	No	$P(c = 1\|e = 1)$	$\frac{S_C+S_G-S_C S_G}{S_C+S_G/P(c=1)-S_C S_G}$	↓[a]	—[a]

Note. $\therefore$ stands for *therefore*. The arrows refer to an increase (↑), decrease (↓), or no change (—) in the acceptance of the logical argument.
[a] Denotes the classic effects.

effect (Beller, 2006; Thompson, 1994). Second, increasing S_G should decrease the acceptance of Modus Tollens, $P(c = 0|e = 0)$. The effect is predicted to be small, and the reason is quite complex. Normally once it is known that $e = 0$ then it is very likely that $c = 0$. But the stronger S_G is, then the more likely that the alternative inhibitory causes were present, in which case it is less certain that c must have been absent (Cummins et al., 1991; De Neys et al., 2002, Experiment 2; De Neys et al., 2003b; Markovits & Handley, 2005, Experiment 1; see also Cummins, 1995). Third, increasing S_I should increase the acceptance of Denying the Antecedent, $P(e = 0|c = 0$; Beller, 2006; De Neys et al., 2002, Experiment 2).

Our goal here is to point out how causal network models may be useful for explaining conditional reasoning effects, with the benefits of a formal yet flexible framework. The research is still insufficient to provide a strong argument for or against the value of this interpretation.

In a similar vein, Ali, Chater, and Oaksford (2011) have used a causal network framework to model conditional reasoning, comparing common cause versus common effect structures. For example, one of the common effect scenarios had two conditionals: "If I do not clean my teeth, then I get cavities" and "If I eat lots of sugar, then I get cavities." Then participants were asked "I got a cavity . . . how likely is it that I did not clean my teeth?" $P(c_1 = 1|e = 1)$ and "I got a cavity and I ate lots of sugar . . . how likely is it that I did not clean my teeth?" $P(c_1 = 1|e = 1, c_2 = 1)$. They found some of the effects predicted by causal structures such as discounting, but they also found some effects that are inconsistent with causal structures, such as violations of the Markov Assumption, $P(c_1 = 1) > P(c_1 = 1|c_2 = 1)$.

In sum, there are some intriguing applications of causal networks to model the acceptability of logical arguments and conditional reasoning. Although these approaches show promise, these paradigms rely heavily upon the application of knowledge that people have about the causal relationships as well as linguistic pragmatics, which pose challenges for assessing the causal structure framework.

Diamond Structures

Diamond structures are unique in that there are two routes from the cause to the effect, and both routes must be simultaneously considered when performing inference. We already discussed a structure with two routes in the section on discounting when the two causes are correlated. Here we use M_1 and M_2 to refer to alternative mediators of the two routes (see Figure 19).

Reasoning About Both Routes Simultaneously

Meder et al. (2008) investigated whether people take M_2 into account when inferring $P(E|M_1)$. To test this, they compared two inferences, when M_1 is observed to be present, $P(e = 1|m_1 = 1)$, versus when one intervenes and sets M_1 to be present, $P(e = 1|\text{set } m_1 = 1)$. If M_1 is observed to be present, then C and M_2 are probably present, so $P(e = 1|m_1 = 1)$ should be very high. In contrast, when M_1 is intervened upon and set to 1, the intervener has no knowledge of the state of C or M_2; the best estimate of M_2 is its base rate. Thus, $P(e = 1|\text{set } m_1 = 1)$ should be lower than $P(e = 1|m_1 = 1)$. Through the same logic, the opposite pattern holds for observing versus setting $m_1 = 0$. In sum, the following asymmetries should hold: $P(e = 1|m_1 = 0) < P(e = 1|\text{set } m_1 = 0) < P(e = 1|\text{set } m_1 = 1) < P(e = 1|m_1 = 1)$.

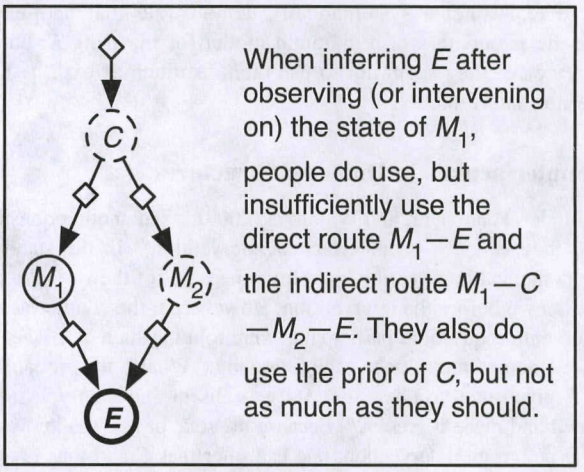


Figure 19. Summary of the $P(E|M_1)$ inference. See Figure 4 for notational conventions.

Meder et al. (2008; Experiment 1) told participants about the diamond structure, and participants experienced a series of trials to learn the parameters. Remarkably, participants' answers to the inference questions reflected the predicted asymmetries. These results suggest that their participants understood the difference between interventions and observations, understood that M_1 and M_2 would be correlated for observations but not for interventions, *and* used both M_1 and M_2 to infer E. In follow-up experiments, Meder, Hagmayer, and Waldmann (2009) also demonstrated that people's inferences are sensitive to the base rate of C and the strength of the causal relations.

Even though these studies did demonstrate the basic normative patterns, the inferences tended to be weaker (i.e., closer to the middle of the scale) than expected. For example, in Meder et al. (2008, Experiment 1), the difference between $P(e = 1|m_1 = 1)$ and $P(e = 1|m_1 = 0)$ and should have been .78 (when converted to a probability scale). However, participants inferred a difference of only .37. This could have been due to underweighting the strength of M_1, or underweighting any of the three other causal strengths (see Figure 19). Additionally, the difference between $P(e = 1|\text{set } m_1 = 1)$ and $P(e = 1|\text{set } m_1 = 0)$ should have been .48, but participants inferred a difference of only .15. This reflects an underweighting of the impact of M_1 on E (see Figure 19).

Similar effects obtained in the 2009 study as well. Meder et al. (2009, Experiment 2) examined how differences in the base rate of C would affect inferences of $P(e = 1|\text{set } m_1 = 0)$. When M_1 is intervened upon and set to 0, the only possible cause of E is M_2, and the probability of E should be higher to the extent that the base rate of C is higher. The manipulation of $P(c = 1)$ did produce a difference in the judgment of E, but the difference was only about half as large as it should be (.15 vs. .30). In sum,

these experiments systematically demonstrate that people do use the parameters of a diamond model for inferring E, but in every case, they seem not to use them as much as expected by the normative model.

Counterfactuals in Diamond Structures

Meder, Hagmayer, and Waldman (2009) asked another question that they called a "counterfactual intervention." In the standard "hypothetical intervention" question, the states of the variables are not known before the intervention. However, in the counterfactual intervention question, participants were told the state of M_1 *before* it was manipulated, such as the following: "What is the probability of E given that you saw that M_1 was absent and then you intervened and made it present?" Because the state of M_1 was known to be 0 before the intervention, one can infer that C and thus M_2 are probably also 0. In this way, the counterfactual intervention question requires reasoning about both routes (M_1-E) and (M_1-C-M_2-E).

Meder, Hagmayer, and Waldman (2009) found that people's inferences were only minimally different comparing standard intervention questions and counterfactual interventions. This lack of a difference could be interpreted as underweighting any or all of the causal strengths along this route or just general confusion about the question.

Intervening on Causal Structures to Produce Desired Outcomes

So far our discussion has focused on inference for its own sake. But, inferences also serve another purpose: They can help us identify interventions that produce desired outcomes (Meder et al., 2010; Sloman & Hagmayer, 2006). We can expand the standard causal network framework introduced in the introduction with utility nodes to represent the desirability of various events. In fact, the "Profit from Tomatoes" node in our Farming Scenario is essentially a utility node. Rationally, it would make sense to choose interventions that maximize the utility over all the utility nodes in the network.

Choosing an intervention to maximize the utility nodes of a causal network requires two steps in addition to performing inferences. First, instead of inferring the value of one node given an intervention, one must infer the value of all the utility nodes for a given intervention and sum across them. Second, one must choose the intervention to maximize expected utility. This decision may seem trivial, but given that people often exhibit probability matching instead of maximization in choice paradigms, it is possible that they will fail to maximize the utility of the network (Eberhardt & Danks, 2011).

Nichols and Danks (2007, Experiment 1) taught people a common effect structure $C_1 \rightarrow E \leftarrow C_2$, in which C_1 was stronger than C_2. Participants could intervene on either C_1 or C_2 to try to produce the E, which was tied to a monetary reward. Not surprisingly, they were more likely to intervene on C_1. Out of the participants who intervened on C_2, most of them incorrectly believed that C_2 was stronger than C_1.

In a second experiment, Nichols and Danks taught participants about a chain structure $C \rightarrow M \rightarrow E$. Intervening on M was more likely to produce E than intervening on C; however, the "cost" of intervening on M was greater than the "cost" of intervening on C, making the average expected payoff higher for C than M. Seventy-eight percent of participants intervened on the variable that, according to their beliefs about the network, would maximize their expected payoff. However, 15% of participants still chose interventions that did *not* maximize expected payoff, according to their own beliefs about the causal structure.

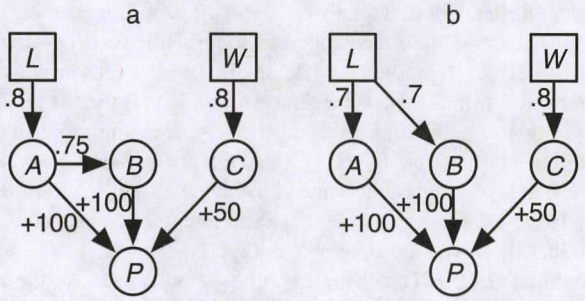


Figure 20. Choosing actions to maximize payoff P.

Hagmayer and Meder (2008, 2012; Meder & Hagmayer, 2009) investigated a similar phenomenon with the structures in Figure 20. The square nodes represent possible interventions, the P node represents an outcome to be maximized, and the plus signs denote the size of the outcome given that a given combination of nodes (A, and or B, and or C) is active. In Hagmayer and Meder's (2012; Experiment 3) study, participants first learned the causal structures (either Figure 20a or 20b) by activating L or W 100 times and observing whether A, B, or C became active and the value of P. Afterward, participants were told that the A node was removed from the network, and they had 10 opportunities to activate L or W in order to maximize P.

Those who believed the structure to be the one in Figure 20a almost always chose W; intervening on L would have no chance of producing P now that A was removed from the network. However, participants who believed the structure to be the one in Figure 20b only chose L 55% of the time, even though they understood that L still had a higher expected value than W. In sum, people's beliefs in the causal structure did have a large influence on their choices, but they also did not choose interventions that would fully maximize the outcome according to their own beliefs. Probability matching is a likely explanation.

These initial studies suggest that, for the most part, people use their beliefs about causal structures to choose actions that will increase payoffs. We speculate that when people confidently believe in a causal structure, they tend to maximize instead of probability match (e.g., Taylor, Landy, & Ross, 2012). But, in the real world, where people often choose interventions with incomplete knowledge of the relevant causal system, the probability matching habit emerges.

General Discussion

In this review, we focused on studies in which people are given a causal structure, learn the parameters, and then make inferences. We started with a review of behavioral studies that examined violations to the Markov Assumption. We then catalogued various inferences that can be made on chain, common cause, one link, common effect, and diamond structures. Finally, we discussed how

people decide to intervene on causal structures in order to produce a desired effect. In this General Discussion, we first discuss the uses of a normative rational analysis. Then we summarize and reorganize key results into two sections based on (a) violations of the Markov Assumption and (b) conservative inferences or underuse of parameter information. We also discuss possible approaches to a more descriptive, semirational model.

The Uses of the Normative, Rational Analysis

The present review was motivated by the relatively recent invention of a normative model for representing and calculating the implications of a system of causal relationships. However, there is no consensus on the value of normative, optimal, rational models in the behavioral sciences. Indeed, the very nature of a normative analysis is a matter of dispute. So, we provide a brief discussion that clarifies our own views on this knotty bundle of questions.

What is an optimal, rational model? Scientific, mathematical, and philosophical analyses produce models of real world situations that can be used to guide actions to achieve goals in an optimally efficient manner. In some cases, the goals are implicit (e.g., logical truth-maintaining coherence; precise forecasts of the operations of a mechanical system), and in others, the goals are stated as part of the model of the situation (e.g., to trade-off expected risk and returns at a designated rate in an investment). Such normative models are evaluated with reference to their accuracy or their usefulness in achieving outcomes in objective physical, biological, or social realities. Some examples of normative models that have been used in the behavioral sciences are elementary logic, probability and other mathematical theories, utility theories and Game Theory in the von Neumann-Morgenstern tradition, "ideal" models for identifying sensory stimulus events, and physics laws of mechanics. For all of these normative models there is close to unanimous consensus among experts that the models are accurate descriptions of the relevant domains of reality. The Bayesian Causal Networks framework is a new candidate for a normative model to represent objective causal systems.

Normative models can be contrasted with descriptive or psychological models that attempt to explain and predict behavior by proposing psychological mechanisms. Some theorists believe that normative models are closely related to psychological-descriptive models (e.g., many economists assume that a "rational man" model provides a good description of the actual behavior of economic agents; many behavioral ecologists believe that optimal models are the best descriptive models for the behavior of foraging animals; cf. Krebs & Davies, 1993). But, most psychologists believe that there are significant differences between the predictions of normative models and actual behavior.

The first conceptual challenge facing behavioral researchers who want to use normative models is to specify the application of a normative framework to a behavioral task. In many cases the identification of an optimal model is not obvious, so alternate rational models must be entertained (see disputes in M. Jones & Love, 2011, and discussion in Holyoak & Cheng, 2011). Examples in the present context include maximizing the total payoff versus maximizing the probability of a payoff when choosing an intervention (e.g., Nichols & Danks, 2007), whether people interpret the scenarios used in typical experiments to be atemporal or temporal (Rottman & Keil, 2012), and whether people intuitively believe that causes combine using noisy-OR or some other function. Even in the highly constrained environment of a psychology experiment, there is always the potential for ambiguity about the scenario, task, goals, and relevant prior knowledge. In sum, claiming a model as optimal or rational for a particular task requires justification and often requires making simplifying assumptions about the task.

The special difficulty of defending a normative model for causality. Identifying a normative framework for causal reasoning is particularly challenging because of its rich and diverse nature. We talk (and think) fluently about many different domains of causality including biological, mechanical, psychological, and social causation: "The fruitworm infestation caused the poor tomato harvest"; "the icy highway caused the traffic accident"; "Jill's intelligence caused her to get a perfect score on the SAT test." We can comprehend the meaning of causal statements despite a lack of understanding as to how they occurred (e.g., "God caused the Red Sea to part"; "Fossil fuel emissions cause global warming," "Smoking causes lung cancer"). We think about causal processes that unfold at many different time frames and orders of magnitude, and we fluently reason about both single cause-effect instances and statistical regularities.

Many find the Causal Networks formalism to be a useful normative framework of objective causation. However, there is still much controversy about using Causal Networks as a foundation for conceptualizing causation. First, there is less acceptance of its status as a normative model than for other popular normative systems (e.g., elementary mathematics, logic, probability, and mechanics). Second, Causal Networks are only a couple of decades old and still changing at a higher rate than older, more established normative systems. Third, there is more disagreement on metaphysical assumptions concerning objective causation than there is on referents of the other normative systems.

Uses of a normative analysis with no claims about its psychologically descriptive validity. Several useful applications of normative models involve no claims about relationships between the normative and psychological-descriptive theories (cf. Garner, 1974, pp. 192–193). Normative frameworks provide a language to describe experimental tasks and goals, to specify at least one procedure for performing a task, and to determine standards for accurate or optimal performance. For example, Morris and Larrick's (1995) analysis of discounting provided a language to discuss discounting [as the relationship between $P(c_1 = 1)$, $P(c_1 = 1|e = 1)$, and $P(c_1 = 1|e = 1, c_2 = 1)$]. Their analysis also clarified the "objectives" that were underspecified in the previous attribution theory literature; depending on the causal structure and parameters, $P(c_1 = 1|e = 1, c_2 = 1)$ should sometimes be greater than, equal to, or less than $P(c_1 = 1)$. Differences between the normative "answers" and human performance are often consequential. Knowing when humans are nonoptimal may be useful in practical endeavors and in guiding the design of remedial procedures. In the case of the present review, we believe it is important to know what kinds of errors people are likely to make when they reason intuitively or analytically, even in a controlled experimental setting, about what's causing what and how to use causal knowledge to bring about desired outcomes.

A closely related approach is to keep normative and descriptive accounts separate, but to pursue a research program to map the two levels onto each other. The most commonly cited inspiration for this research tactic is David Marr's (1982) three-level framework, which distinguished between a Computational Level (a functional analysis, often a normative model, including the actor's goals), an Algorithmic-Representational Level (the descriptive-psychological model), and an Implementational Level (a neural-biological model)—and which promoted formal mappings between adjacent levels.

Normatively inspired descriptive models. Many behavioral researchers go a step further and use normative models as an inspiration for psychological-descriptive theories or principles (see J. R. Anderson, 1990; J. R. Anderson & Milson, 1989, as exemplars). They first complete a normative analysis of the task and then use that analysis (with samples of behavioral data) to guide the invention of a descriptive model. The most commonly mentioned justification for this interaction between the two types of models is to note that humans are selected by evolution and shaped by learning to excel at tasks that are important to our survival, so that many of the normative principles are likely to be "wired-in" genetically or learned from individual experience as an adaptive strategy.

When applying a rational framework to empirical results, it is often found that human minds are bounded or lazy in ways that prevent them from performing the optimal calculations required for "full rationality" (Gigerenzer, Todd, & The ABC Research Group, 1999; Kahneman, 2003; Payne, Bettman, & Johnson, 1993; Shah & Oppenheimer, 2008; Simon, 1955). The notion that informal causal inference would follow shortcuts is especially plausible when one thinks though all of the calculations that would be necessary for a sufficient model of the optimal computation (Fernbach & Rehder, 2012; or see the complex equations in this article). Because the application of Causal Network models is so new, there are no full-fledged general proposals for the manner in which the rational model should be adjusted to be more descriptive. In the following sections, we cite some proposals for parts of the problem.

The normative model is the descriptive model. The most extreme approach is to say we don't need a descriptive model because we can predict behavior from only the normative model. No one so far has explicitly proposed this claim for causal reasoning, although some researchers have come close, by emphasizing the correspondences between Bayesian networks and participants' judgments (e.g., Krynski & Tenenbaum, 2007; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). Yet we believe that most researchers expect there will be some reliable differences between normative and descriptive accounts (M. Jones & Love, 2011, and commentary). Our review refutes the strong claim with several examples of consistent discrepancies between human judgments and the implications of well-defined causal networks.

In the next sections, we discuss the two main deviations from the normative model: violations of the Markov Assumption and conservative or weak inferences. We also discuss possible modifications to the normative model to make it more descriptive.

Summary of Main Results

The studies we have reviewed almost all had the goal of examining whether an experimental manipulation produced a significant effect in the direction predicted by the normative model. We also looked for patterns of the quantitative fit of the normative model, such as conservative biases where the inference was "in the right direction" but was too weak. We emphasized systematic patterns across experiments rather than deviations in particular means in single experiments. However, because many of these inferences have been studied in only one or two experiments and often those experiments were not designed to investigate the particular comparison we were interested in, some of our conclusions are educated judgment calls. We discuss these findings in terms of the farming example used in the introduction, reprinted in Figure 21.

Violations of the Markov assumption. The Markov Assumption specifies which nodes should be ignored for a particular inference, which simplifies reasoning. However, many studies found violations of the Markov Assumption. For example, if one knows that there was a poor tomato harvest (T), learning about an early frost (F) should not have any impact on inferences about profit (P), yet it did. Likewise, if one knows that there was an early frost on the farm (F), learning that there was a poor cantaloupe harvest or a good cantaloupe harvest (C) should not have any bearing on whether there was a poor tomato harvest (T). Yet it did here, too. Burnett (2004) also found bigger violations for closer variables (e.g., F would have a bigger effect than C on inferring P even when the state of T is known).

Some of these violations can be explained through alternative accounts that justify the apparent deviation with a rational or adaptive interpretation such as imagining additional nodes in the network or additional causal relationships outside those specified by the experimenter (e.g., Burnett, 2004). Back to the farming example, perhaps observing that there is a poor cantaloupe harvest is a sign that there was not enough rain, a variable not represented in the network, which might also cause a poor tomato harvest. Everyday causal systems are more complex than those in the experiments. Because of this complexity, some skeptics of the Causal Networks approach for engineering and data mining have argued that the Markov Assumption is unrealistically restrictive (Cartwright, 1999, 2001, 2002).

A more philosophical justification derives from the probabilistic nature of causality in these experiments. When a cause occurs and an effect does not (or vice versa), one interpretation implies that there must be an additional (generative or inhibitory) cause(s) that also influences the effect (Rottman, Ahn, & Luhmann, 2011). More fundamentally, if people act as if we live in a Laplacean world (i.e., if we know the state of everything in the universe then it is possible to perfectly predict the future), any contradiction between the causes and the predicted effects implies that there

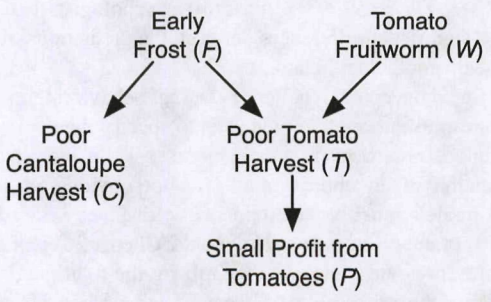


Figure 21. Farming scenario.

must be unknown factors. A person who conceives of causal relationships in this manner would certainly interpret an experimenter's description of a small set of probabilistically related events as a subset of all relevant events. Believing that there are unobserved causes is not a problem for the causal network framework per se. But it is a problem if these unobserved relationships result in additional correlations between observed variables. Admitting this possibility undermines the validity of the experimental tests of the causal network framework, and it also challenges the validity of the framework in all applications.

Our view is that there are enough violations of the Markov independence condition, in cases where "importing" additional causal links was highly implausible or unjustified, to force the conclusion that humans reliably violate the principle. As noted before, it is also informative that these additional correlations have always been found to be positive; there is no reason a priori why they would not be negative.

There are several potential explanations for these patterns of reasoning. First, some people may engage in associative reasoning (e.g., Rehder, 2012). Associative style reasoning implies that people don't distinguish the direction of causal relationships (such as the difference between a common cause and common effect structure). Processes like second-order conditioning could potentially explain why people think that screened-off variables are still relevant. Alternatively, Hagmayer and Waldmann (2002) developed a constraint-satisfaction model of causal learning and reasoning. A characteristic of this model is that it is easy to learn individual causal relationships but harder to understand entire causal structures and the conditional and unconditional independencies (e.g., the difference between common cause vs. common effect structures). A related approach is to propose reasoning "locally" on subsets of the graph or single causal relations at a time (e.g., Fernbach & Sloman, 2009; Kruschke, 2006; Waldmann et al., 2008). In sum, these persistent violations warrant considering nonnormative consistency-seeking explanations.

Conservative inferences. Another result that has been reported in many different studies is that people made less extreme inferences than are implied by the parameters of the causal networks. "Base rate neglect" is the most obvious example of an undersensitive inference. Consider the one-link structure: *early frost → poor cantaloupe harvest*. One would expect the probability of an early frost given a poor cantaloupe harvest to be higher than the prior probability of an early frost, although this was not always observed (Meder, Hagmayer, & Waldmann, 2009).

Consider the chain structure: *early frost → poor tomato harvest → small profit from tomatoes*. What is the chance of a small profit given an early frost? For analogous questions, Baetu and Baker (2009) found that transitive inferences are not as strong as they should be. Rehder and Kim (2010) asked their participants to infer the marginal probability of small profit from tomatoes. Although participants' inferences were influenced by the appropriate parameters (the base rate of early frost and the strengths of the causal links), they were not as sensitive as they should have been.

Consider the common cause structure: *poor cantaloupe harvest ← early frost → poor tomato harvest*. During years in which there is a poor (vs. good) cantaloupe harvest, it is likely that there would also be a poor (good) tomato harvest. In analogous situations in which people separately learned about the two causal relationships, they did not fully understand the extent to which effects of a common cause were correlated (Hagmayer & Waldmann, 2000; Perales et al., 2004).

Consider the common effect structure: *early frost → poor tomato harvest ← tomato fruitworm infestation*. Learning that there was a poor tomato harvest makes an infestation more likely, but subsequently learning that there was an early frost suggests that there *was not* an infestation; the frost "explains away" the poor tomato harvest. Although "explaining away" is considered to be a hallmark of causal reasoning, the existing research has found it to be weaker than it should be, if present at all (Morris & Larrick, 1995; Rehder, 2012; Sussman & Oppenheimer, 2011). Fernbach et al. (2011) asked participants questions analogous to "An early frost occurred; what is the probability that there was a poor tomato harvest?" Participants tended to ignore the possibility that an infestation could also cause a poor tomato harvest.

Because Figure 21 does not have a diamond, we modified it (see Figure 22). Meder et al. (2008) and Meder, Hagmayer, and Waldmann (2009) asked participants questions analogous to, "What is the probability of a small total profit given that there is a poor cantaloupe harvest?" implying that there probably was also an early frost and probably also a poor tomato harvest. They also asked the same question, "given that the cantaloupes were poisoned?" which implies nothing about an early frost or the tomato harvest. Both of these inferences were closer to the middle of the scale than expected, which could reflect insufficient use of the parameters.

There are a number of possible explanations for conservative inferences that derive from characteristics of the experimental tasks. First, it is possible that even though participants in these experiments were told the causal structure, they did not accept the experimenter's statement of the causal structure. If people are uncertain about the causal structure they might perform inferences over multiple possible structures (Meder, Mayrhofer, & Waldmann, 2009; see Schum & Martin, 1982, for a related problem in law). However, many of the studies we review used novel variables, and it is not clear why participants would have rejected the experimenters' cover stories about the causal structure, especially when the learning data also matched the causal structure.

Second, it is possible that people had not fully learned the parameters of the causal model; if they had observed more evidence, their beliefs in the parameters might have been stronger. Meder, Hagmayer, and Waldmann (2009) proposed that their participants' parameter estimates might have been influenced by a prior distribution

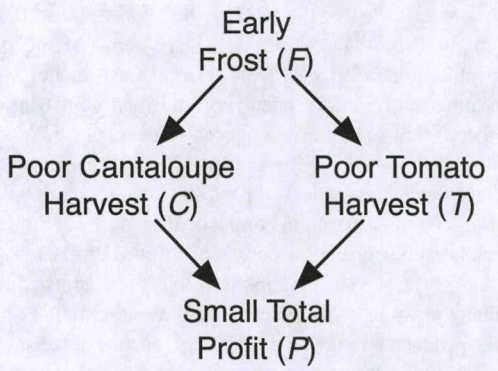


Figure 22. Diamond farming scenario.

(e.g., a uniform prior) that could have pulled inferences toward the middle of the scale. However, other theorists have argued that participants have nonuniform priors in mind. For example, Lu et al. (2008; see also Yeung & Griffiths, 2011) suggested that people expect causes to be either strong or nonexistent, but not moderately strong. If people actually used these priors, then their inferences would be *more extreme* than the standard analysis; yet, people's inferences tend to be *more conservative*.

There are two other pieces of evidence suggesting that conservative inferences are not just due to insufficient or preasymptotic learning of the parameters or to averaging over participants with extreme but different judgments. First, the standard observation of base rate neglect (e.g., predictive value of breast cancer given a positive mammogram $P(c = 1|e = 1)$; Eddy, 1982) occurs even when people are explicitly told the parameters. Indeed, base rate neglect has traditionally been found to be more extreme in situations in which the base rates are explicitly stated, compared to when they are learned from experience (Christensen-Szalanski & Beach, 1982; Koehler, 1996). Second, some of the studies found conservative inferences even compared to participants' own stated beliefs in the parameters (e.g., Fernbach et al., 2011; Morris & Larrick, 1995). Thus, we conclude that conservative inferences are caused by something more than insufficient learning of the parameters.

Third, it is difficult to separate true conservative reasoning from methodological artifacts associated with the rating scales used in all of the studies that ask for numerical ratings. It is plausible that some of the conservative habits are merely response biases produced by using response formats with a salient, "safe" or "compromise" midpoint. This artifact cannot be evaluated without systematic variations of the response scale formats, tests on inferences that involve different regions on the scales, and performance-contingent incentives.

Overall, some of the conservatism in judgments is likely due to general habits of caution. But, we also believe that there are hints in the conservative patterns of inferences that additional judgment habits are involved. We think it is unlikely that deliberate reasoning processes exactly map onto the Bayesian calculations. We conjecture that anchor and insufficient adjustment habits are plausible psychologically (cf. Lopes, 1987). The problem with this interpretation is that it simply relabels the observed results, without providing deeper understanding, unless the anchoring process is further specified.

Let's walk through a speculation on anchoring strategies. Consider inferences on the C→E structure. Suppose the following causal parameters are provided via verbal-numerical instructions [$P(c = 1) = .30$, $P(e = 1|c = 1) = .80$, $P(e = 1|c = 0) = .40$], which imply $P(e = 1) = .52$. What are some of the plausible anchor values for inferring $P(e = 1)$? (a) One anchor would be zero; assume that E is not occurring and then adjust upward for causal forces that increase its chances of occurring [$P(e = 1|c = 1) = .80$ and $P(e = 1|c = 0) = .40$]. (b) Another anchor could be the salient value $P(e = 1|c = 1) = .80$; then adjust down toward $P(c = 1|e = 0) = .40$, or in the opposite direction. (c) Some people might anchor on a midpoint between $P(e = 1|c = 0) = .40$ and $P(e = 1|c = 1) = .80$, perhaps .60, and then adjust downward given that $P(c = 1) = .30$. Note that these alternative anchoring strategies produce a range of predictions: anchor on zero, which is likely to produce a low rating, versus anchor on .80, which is likely to produce a high rating. The predictions are blurred further by the plausible assumption that different participants are likely to anchor on different parameters.

We can also speculate about psychological processes when the causal structure is learned from samples rather than declaratively through words and numbers. For an inference like $P(e = 1)$, participants might assess the memory strength or frequency in memory of $(e = 1)$ experiences, in which case the assessment is likely to be regressive with overestimated low frequencies and underestimated high frequencies (Attneave, 1953; Zacks & Hasher, 2002). For an inference like $P(c = 1|e = 1)$, participants could try to recall the percentage of $(c = 1)$ experiences out of the recalled set of $(e = 1)$ experiences.

This discussion makes it obvious that anyone who wants to make an empirical argument for an alternative to the Bayesian calculation will have to be clear about the alternative calculations that are proposed and increase the control and precision of the experimental methods. In fact, we hope researchers proceed in this fashion, as we do *not* believe that the humans' explicit inferences about causal relationships are fully Bayesian. We also believe that some kind of serial averaging process is the most likely candidate for an alternative calculation, given the vast number of averaging results in the judgment literature and given that averaging, for most parameter values in the research we have reviewed, produces conservative final estimates.

We should also note a final complexity. One important aspect of the weak causal inferences is that they seem to run against the Markov violations. Take the structure $C \rightarrow M \rightarrow E$. A typical violation of the Markov assumption involves inferring $P(e = 1|m = 1, c = 1) > P(e = 1|m = 1, c = 0)$. The two judgments are *too far apart*, when they should be equal; C affects the inference about E when it should have been "screened off" by the knowledge of the mediator (M). In contrast, a standard too-weak transitive inference involves inferring that $P(e = 1|c = 1)$ and $P(e = 1|c = 0)$ are *too close together*. The only difference between these two sets of findings is whether the state of M is known or not. Recall that some researchers (e.g., Rehder & Burnett, 2005) proposed adding a "hidden mechanism" node to explain the Markov Violations, but adding such a node would lead to overly strong rather than weak transitive inference. The implication is that it is doubtful that there is a unitary rational explanation for these two results.

A potential way to model these two findings is with a linear averaging approach. When inferring E, M gets most of the weight, but C still gets some weight. This approach could potentially capture the fact that C is weighted too little for transitive inferences, but it is weighted too much (it should have zero weight) when M is known. This approach might also be useful for explaining how people infer M on the chain $C \rightarrow M \rightarrow E$ or C on the common cause $E_1 \leftarrow C \rightarrow E_2$. There is not much research on how normatively people make judgments like $P(m = 1|c = 1, e = 1)$, but it is likely that people use some sort of linear averaging instead of a Bayesian likelihood ratio calculation (e.g., N. H. Anderson, 1996; Lopes, 1987).

Summary of Possible Psychological Processes Involved in Causal Inference

Here we summarize some of the judgment problems faced in causal inference and present some potential cognitive process explanations; references appear in sections above.

Causal structures are complex. People may have difficulty understanding all the dependencies and conditional independencies implied by structures with multiple variables even if they understand each of the individual links. For example, explaining away and the independencies implied by the Markov assumption are not necessarily intuitive. Constraint satisfaction and associative reasoning strategies may provide some people with alternative representations for the structures. "Local" reasoning on parts of the structure could also explain why people have difficulty understand properties of the structure that emerge when reasoning about three or more nodes simultaneously.

Too much information and integration is confusing. Performing the full Bayesian calculations requires reasoning about many nodes simultaneously, understanding how causes combine in complex ways (e.g., noisy-OR rule), and understanding how to use multiple parameters for a single inference. Even though anchoring is more of a description than a process model, it suggests a way to reduce complexity by focusing primarily on one piece of information and then sequentially adjusting for other pieces of information.

Too much uncertainty. When one is uncertain about the causal structure or strengths, one might use "safe" defaults for judgments, such as the middle of the scale, or potentially rely on base rates with little updating. Uncertainty can also be built into the normative framework by integrating over possible structures or conditioning on sample size.

Limited memory. When one experiences the probabilistic relationships between multiple variables, the number of cells in the joint probability table (e.g., Table 2) required to represent those experiences becomes very large. Focusing on the parameters instead of the contingencies simplifies the reasoning process, although we do not know whether people naturally reason using the parameters or the raw experiences. Either way, memory biases could impact the assessment of parameters or judgments based directly on a mental version of the joint probabilities.

In sum, there are a variety of potential cognitive strategies and biases that could affect inferences on causal structures. We hope that summarizing these possibilities will encourage future research.

Conclusions

The Bayesian Probabilistic Causal Networks framework has stimulated a productive research program on human inferences on causal networks. Such inferences have clear analogues in everyday judgments about social attributions, medical diagnosis and treatment, legal reasoning, and in many other domains involving causal cognition. So far, research suggests two persistent deviations from the normative model. People's inferences of one event are often inappropriately influenced by other events that are normatively irrelevant; they are unconditionally independent or are "screened off" by intervening nodes. At the same time, people's inferences tend to be weaker than are warranted by the normative framework.

These conclusions do not sharply constrain the form of a descriptive model for causal reasoning. At one end of the spectrum, some psychologists may want to ignore the normative framework (although we hope they would still consider its value as a model for objective causation). Such a theorist might want to "work up" from the lower implementational level, such as associative networks or constraint satisfaction networks, which can mimic many of the properties of normative Causal Networks but are not committed to the strict normative calculus.

Another option is to start with the normative Causal Networks and to relax some of the assumptions. Some candidates for "relaxation" include (a) shifting from exhaustive hypothesis spaces to attention-limited subsets of cognitively salient hypotheses, (b) considering alternative prior belief probability distributions (e.g., Lu et al., 2008), (c) limiting updating inferences to a subset of network nodes (presumably because of working memory limits, attention limits, pragmatics, or proximity; e.g., Burnett, 2004), (d) conditioning confidence in experimentally learned parameter values on sample size or credibility to more realistically represent uncertainty about the network (cf. Winkler & Murphy, 1973), and (e) experimentally verifying that the participants in experiments have not added plausible nodes or links to the experimenter-defined causal system (e.g., Burnett, 2004).

Causal reasoning is one dramatic example of an exceptionally sophisticated system of inferences that approximates many properties of normative belief systems. The research we reviewed has shown that when the normative calculations of causal networks imply that the probability of an event should increase, the judgments usually go up; when they imply a decrease, judgments usually go down. At the same time, the experimental literature contains some substantial and systematic discrepancies between human inferences and those of the normative Causal Network framework. Empirical and theoretical research on these discrepancies is an important frontier for our exploration of human cognition and human nature more generally.

References

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional inference: Causal models or mental models. *Cognition, 119,* 403–418. doi:10.1016/j.cognition.2011.02.005

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96,* 703–719. doi:10.1037/0033-295X.96.4.703

Anderson, N. H. (1996). *A functional theory of cognition.* Mahwah, NJ: Erlbaum.

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology, 46,* 81–86. doi: 10.1037/h0057955

Baetu, I., & Baker, A. G. (2009). Human judgments of positive and negative causal chains. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 153–168. doi:10.1037/a0013764

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgment. *Acta Psychologica, 44,* 211–233. doi:10.1016/0001-6918(80)90046-3

Beller, S. (2006). What we can learn from causal conditional reasoning about the naive understanding of causality. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 59–64). Mahwah, NJ: Erlbaum.

Bennett, J. (2003). *A philosophical guide to counterfactuals.* Oxford, England: Oxford University Press. doi:10.1093/0199258872.001.0001

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311,* 1020–1022. doi:10.1126/science.1121872

Buchanan, D. W., & Sobel, D. M. (2011). Children posit hidden causes to

explain causal variability. L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 3098–3103). Austin, TX: Cognitive Science Society.

Burnett, R. C. (2004). *Inference from complex causal models* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3156566)

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139167093

Cartwright, N. (2001). What is wrong with Bayes nets? *Monist, 84,* 242–264. doi:10.5840/monist20018429

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two. *British Journal for the Philosophy of Science, 53,* 411–453. doi:10.1093/bjps/53.3.411

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine, 12*(4), 50–63.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405. doi:10.1037/0033-295X.104.2.367

Christensen-Szalanski, J. J., & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance, 29,* 270–278. doi:10.1016/0030-5073(82)90260-4

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition, 23,* 646–658. doi:10.3758/BF03197265

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19,* 274–282. doi:10.3758/BF03211151

Danks, D. (2009). The psychology of causal perception and reasoning. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation* (pp. 447–470). Oxford, England: Oxford University Press. doi:10.1093/oxfordhb/9780199279739.003.0022

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition, 30,* 908–920. doi:10.3758/BF03195776

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology, 15,* 161–176. doi:10.1080/09541440244000058

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition, 31,* 581–595. doi:10.3758/BF03196099

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgment of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 36,* 29–50.

Eberhardt, F., & Danks, D. (2011). Confirmation in the cognitive sciences: The problematic case of Bayesian models. *Minds and Machines, 21,* 389–410. doi:10.1007/s11023-011-9241-3

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, England: Cambridge University Press.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 321–335. doi:10.1037/0278-7393.29.2.321

Evans, J. St. B. T., & Over, D. E. (2004). *If.* Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780198525134.001.0001

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General, 140,* 168–185. doi:10.1037/a0022100

Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. doi:10.1037/a0031851

Fernbach, P. M., & Rehder, B. (2013). Cognitive shortcuts in causal inference. *Argument & Computation, 4,* 64–88. doi: 10.1080/19462166.2012.682655.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 678–693. doi:10.1037/a0014928

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gelman, A., & Meng, X.-L. (Eds.). (2004). *Applied Bayesian modeling and causal inference from incomplete data perspectives.* New York, NY: Wiley. doi:10.1002/0470090456

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704. doi:10.1037/0033-295X.102.4.684

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart.* New York, NY: Oxford University Press.

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

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 3–32. doi:10.1037/0033-295X.111.1.3

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 334–384. doi:10.1016/j.cogpsych.2005.05.004

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116,* 661–716. doi:10.1037/a0017201

Hagmayer, Y., & Meder, B. (2008). Causal learning through repeated decision making. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 179–184). Austin, TX: Cognitive Science Society.

Hagmayer, Y., & Meder, B. (2012). Repeated causal decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. doi:10.1037/a0028643

Hagmayer, Y., & Sloman, S. (2009). Decision makers conceive of their choices as interventions. *Journal of Experimental Psychology: General, 138,* 22–38. doi:10.1037/a0014585

Hagmayer, Y., & Waldmann, M. R. (2000). Simulating causal models: The way to structural sensitivity. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 214–219). Austin, TX: Cognitive Science Society.

Hagmayer, Y., & Waldmann, M. R. (2002). A constraint satisfaction model of causal learning and reasoning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 405–410). Mahwah, NJ: Erlbaum.

Hagmayer, Y., & Waldmann, M. R. (2007). Inferences about unobserved causes in human contingency learning. *Quarterly Journal of Experimental Psychology, 60,* 330–355. doi:10.1080/17470210601002470

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science, 31,* 765–814. doi:10.1080/03640210701530755

Hiddleston, E. (2005). A causal theory of counterfactuals. *Noûs, 39,* 632–657. doi:10.1111/j.0029-4624.2005.00542.x

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology, 62,* 135–163. doi:10.1146/annurev.psych.121208.131634

Jara, E., Vila, J., & Maldonado, A. (2006). Second-order conditioning of human causal learning. *Learning and Motivation, 37,* 230–246. doi:10.1016/j.lmot.2005.12.001

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79,* 1–17. doi:10.1037/h0093874

Jensen, F. J., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs*. New York, NY: Springer-Verlag. doi:10.1007/978-0-387-68282-2

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3,* 1–24. doi:10.1016/0022-1031(67)90034-0

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34,* 169–188. doi:10.1017/S0140525X10003134

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720. doi:10.1037/0003-066X.58.9.697

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454. doi:10.1016/0010-0285(72)90016-3

Khemlani, S. S., & Oppenheimer, D. M. (2010). When one model casts doubt on another: A levels-of-analysis approach to causal discounting. *Psychological Bulletin, 137,* 195–210. doi:10.1037/a0021809

Kim, N. S., Luhmann, C. C., Pierce, M. L., & Ryan, M. M. (2009). The conceptual centrality of causal cycles. *Memory & Cognition, 37,* 744–758. doi:10.3758/MC.37.6.744

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19,* 1–17. doi:10.1017/S0140525X00041157

Krebs, J. R., & Davies, N. B. (1993). *An introduction to behavioural ecology* (4th ed.). Oxford, England: Blackwell.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113,* 677–699. doi:10.1037/0033-295X.113.4.677

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General, 136,* 430–450. doi:10.1037/0096-3445.136.3.430

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 856–876. doi:10.1037/0278-7393.30.4.856

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 451–460. doi:10.1037/0278-7393.32.3.451

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780195176803.003.0011

Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B. Methodological, 50,* 157–224.

Liu, I., Lo, K., & Wu, J. (1996). A probabilistic interpretation of "if–then". *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49,* 828–844.

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica, 64,* 167–185. doi:10.1016/0001-6918(87)90005-9

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115,* 955–984. doi:10.1037/a0013256

Luhmann, C. C., & Ahn, W. (2007). BUCKLE: A model of unobserved cause learning. *Psychological Review, 114,* 657–677. doi:10.1037/0033-295X.114.3.657

Markovits, H., & Handley, S. (2005). Is inferential reasoning just probabilistic reasoning in disguise? *Memory & Cognition, 33,* 1315–1323. doi:10.3758/BF03193231

Marr, D. (1982). *Vision: A computational investigation into human representation and processing of visual information*. San Diego, CA: Freeman.

Mayrhofer, R., Goodman, N. D., Waldmann, M. R., & Tenenbaum, J. B. (2008). Structured correlation from the causal background. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 303–308). Austin, TX: Cognitive Science Society.

Mayrhofer, R., Hagmayer, Y., & Waldmann, M. R. (2010). Agents and causes: A Bayesian error attribution model of causal reasoning. In R. Camtrabone & S. Ohlsson (Eds.), *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Meder, B., Gerstenberg, T., Hagmayer, Y., & Waldmann, M. R. (2010). Observing and intervening: Rational and heuristic models of causal decision making. *The Open Psychology Journal, 3,* 119–135.

Meder, B., & Hagmayer, Y. (2009). Causal induction enables adaptive decision making. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Vol. *70,* pp. 1651–1656). Austin, TX: Cognitive Science Society.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review, 15,* 75–80. doi:10.3758/PBR.15.1.75

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & Cognition, 37,* 249–264. doi:10.3758/MC.37.3.249

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2009). A rational model of elemental diagnostic inference. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2176–2181). Austin, TX: Cognitive Science Society.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review, 102,* 331–355. doi:10.1037/0033-295X.102.2.331

Nichols, W., & Danks, D. (2007). Decision making using learned causal structures. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 1343–1348). Austin, TX: Cognitive Science Society.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111,* 455–485. doi:10.1037/0033-295X.111.2.455

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology Learning, Memory, and Cognition, 26,* 883–899. doi:10.1037/0278-7393.26.4.883

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 680–693. doi:10.1037/0278-7393.29.4.680

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology, 54,* 62–97. doi:10.1016/j.cogpsych.2006.05.002

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781139173933

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

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

Perales, J., Catena, A., & Maldonado, A. (2004). Inferring non-observed correlations from causal scenarios: The role of causal knowledge. *Learning and Motivation, 35,* 115–135. doi:10.1016/S0023-9690(03)00042-0

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68,* B93–B101. doi:10.1016/S0010-0277(98)00053-5

Rehder, B. (2006). *Human deviations from normative causal reasoning.* Poster session presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, British Columbia, Canada.

Rehder, B. (2011). Reasoning with conjunctive causes. In L. Carlson, C. Hőlscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Boston, MA: Cognitive Science Society.

Rehder, B. (2012). *Independence and nonindependence in human causal reasoning.* Manuscript submitted for publication.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50,* 264–314. doi:10.1016/j.cogpsych.2004.09.002

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effect of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360. doi:10.1037/0096-3445.130.3.323

Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1171–1206. doi:10.1037/a0019765

Rehder, B., & Martin, J. B. (2011). A generative model of causal cycles. In L. Carlson, C. Hőlscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Reips, U.-D., & Waldmann, M. R. (2008). When learning order affects sensitivity to base rates. *Experimental Psychology, 55,* 9–22. doi:10.1027/1618-3169.55.1.9

Rips, L. J. (2010). Two causal theories of counterfactual conditionals. *Cognitive Science, 34,* 175–221. doi:10.1111/j.1551-6709.2009.01080.x

Rottman, B. M., Ahn, W., & Luhmann, C. C. (2011). When and how do people reason about unobserved causes? In P. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 150–183). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780199574131.003.0008

Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: Observations and interventions. *Cognitive Psychology, 64,* 93–125. doi:10.1016/j.cogpsych.2011.10.003

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law & Society Review, 17,* 105–152. doi:10.2307/3053534

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134,* 207–222. doi:10.1037/0033-2909.134.2.207

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99–118. doi:10.2307/1884852

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

Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences, 10,* 407–412. doi:10.1016/j.tics.2006.07.001

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science, 29,* 5–39. doi:10.1207/s15516709cog2901_2

Spellman, B. A. (1996). Acting as intiutive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7,* 337–342. doi:10.1111/j.1467-9280.1996.tb00385.x

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search.* New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-2748-9

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27,* 453–489. doi:10.1207/s15516709cog2703_6

Sussman, A. B., & Oppenheimer, D. (2011). A causal model theory of judgment. In C. Hölscher, L. Carlson, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1703–1708). Austin, TX: Cognitive Science Society.

Taylor, E. G., Landy, D. H., & Ross, B. H. (2012). The effect of explanation in simple binary prediction tasks. *The Quarterly Journal of Experimental Psychology, 65,* 1361–1375. doi:10.1080/17470218.2012.656664

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22,* 742–758. doi:10.3758/BF03209259

von Sydow, M., Hagmayer, Y., Meder, B., & Waldmann, M. R. (2010). How causal reasoning can bias empirical evidence. In R. Camtrabone & S. Ohlsson (Eds.), *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society* (Vol. *6,* pp. 2087–2092). Austin, TX: Cognitive Science Society.

von Sydow, M., Meder, B., & Hagmayer, Y. (2009). A transitivity heuristic of probabilistic causal reasoning. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Vol. *1,* pp. 803–808). Amsterdam, the Netherlands: Cognitive Science Society.

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

Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science, 31,* 233–256. doi:10.1080/15326900701221231

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. In N. Chatern & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 453–484). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780199216093.003.0020

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82,* 27–58. doi:10.1016/S0010-0277(01)00141-X

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology Learning, Memory, and Cognition, 31,* 216–227. doi:10.1037/0278-7393.31.2.216

Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1102–1107). Mahwah, NJ: Erlbaum.

Walsh, C. R., & Sloman, S. A. (2004). Revising causal beliefs. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1423–1427). Mahwah, NJ: Erlbaum.

Walsh, C. R., & Sloman, S. A. (2007). Updating beliefs with causal models: Violations of screening off. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A festschrift for Gordon H. Bower (345–358).* New York, NY: Erlbaum.

Winkler, R. L., & Murphy, A. H. (1973). Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance, 10,* 252–270. doi:10.1016/0030-5073(73)90017-2

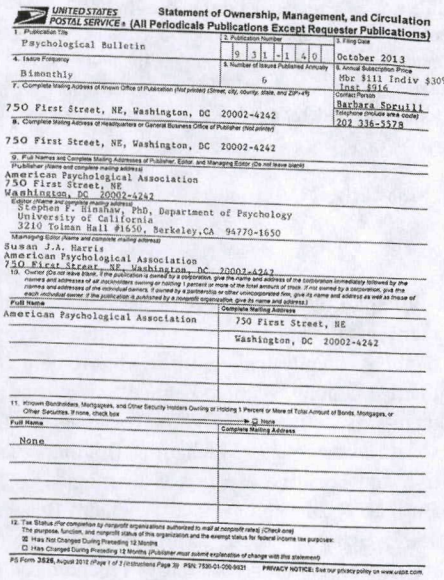
Woodward, J. (2003). *Making things happen: A theory of causal explanation.* New York, NY: Oxford University Press.

Yeung, S., & Griffiths, T. L. (2011). Estimating human priors on causal strength. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1709–1714). Austin, TX: Cognitive Science Society.

Yuille, A. L., & Lu, H. (2008). The noisy-logical distribution and its application to causal inference. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (Vol. 20, pp. 1673–1680). Cambridge, MA: MIT Press.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *Etc.: Frequency processing and cognition* (pp. 21–36). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780198508632.003.0002

Received March 18, 2012
Revision received December 14, 2012
Accepted December 23, 2012 ■



UNITED STATES POSTAL SERVICE® **Statement of Ownership, Management, and Circulation (All Periodicals Publications Except Requester Publications)**

1. Publication Title	2. Publication Number	3. Filing Date
Psychological Bulletin	0 3 1 - 1 4 0	October 2013
4. Issue Frequency	5. Number of Issues Published Annually	6. Annual Subscription Price
Bimonthly	6	Mbr $111 Indiv $309 Inst $916

7. Complete Mailing Address of Known Office of Publication (Not printer) (Street, city, county, state, and ZIP+4®)
750 First Street, NE, Washington, DC 20002-4242
Contact Person: Barbara Spruill
Telephone (Include area code): 202 336-5578

8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer)
750 First Street, NE, Washington, DC 20002-4242

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)
Publisher (Name and complete mailing address)
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242
Editor (Name and complete mailing address)
Stephen F. Hinshaw, PhD, Department of Psychology
University of California
3210 Tolman Hall #1650, Berkeley,CA 94770-1650
Managing Editor (Name and complete mailing address)
Susan J.A. Harris
American Psychological Association
750 First Street, NE, Washington, DC 20002-4242

10. Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

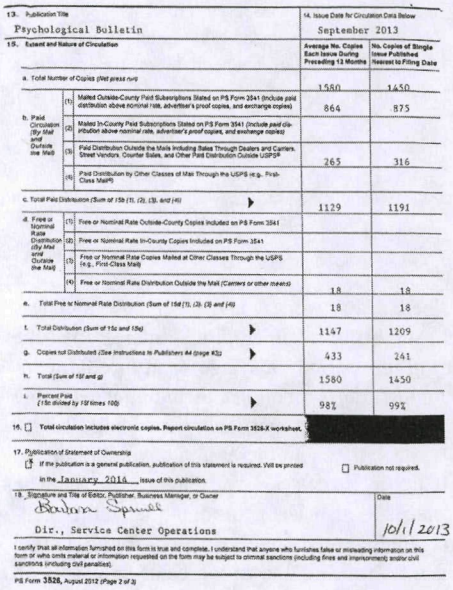
Full Name	Complete Mailing Address
American Psychological Association	750 First Street, NE
	Washington, DC 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address
None	

12. Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526, August 2012 (Page 1 of 3 (Instructions Page 3)) PSN: 7530-01-000-9931 PRIVACY NOTICE: See our privacy policy on www.usps.com



13. Publication Title	14. Issue Date for Circulation Data Below
Psychological Bulletin	September 2013

15. Extent and Nature of Circulation			Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)			1580	1450
b. Paid Circulation (By Mail and Outside the Mail)	(1)	Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	864	875
	(2)	Mailed In-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)		
	(3)	Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	265	316
	(4)	Paid Distribution by Other Classes of Mail Through the USPS (e.g., First-Class Mail®)		
c. Total Paid Distribution (Sum of 15b (1), (2), (3), and (4))			1129	1191
d. Free or Nominal Rate Distribution (By Mail and Outside the Mail)	(1)	Free or Nominal Rate Outside-County Copies included on PS Form 3541		
	(2)	Free or Nominal Rate In-County Copies included on PS Form 3541		
	(3)	Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g., First-Class Mail)		
	(4)	Free or Nominal Rate Distribution Outside the Mail (Carriers or other means)	18	18
e. Total Free or Nominal Rate Distribution (Sum of 15d (1), (2), (3) and (4))			18	18
f. Total Distribution (Sum of 15c and 15e)			1147	1209
g. Copies not Distributed (See Instructions to Publishers #4 (page #3))			433	241
h. Total (Sum of 15f and g)			1580	1450
i. Percent Paid (15c divided by 15f times 100)			98%	99%

16. ☐ Total circulation includes electronic copies. Report circulation on PS Form 3526-X worksheet. [redacted]

17. Publication of Statement of Ownership
☐ If the publication is a general publication, publication of this statement is required. Will be printed in the January 2014 issue of this publication. ☐ Publication not required.

18. Signature and Title of Editor, Publisher, Business Manager, or Owner
Barbara Spruill
Dir., Service Center Operations
Date: 10/1/2013

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

PS Form 3526, August 2012 (Page 2 of 3)