Psychological Bulletin

When One Model Casts Doubt on Another: A Levels-of-Analysis Approach to Causal Discounting

Sangeet S. Khemlani, and Daniel M. Oppenheimer

Online First Publication, December 13, 2010. doi: 10.1037/a0021809

Psychological Bulletin

0033-2909/10/$12.00 DOI: 10.1037/a0021809

When One Model Casts Doubt on Another: A Levels-of-Analysis Approach to Causal Discounting

Sangeet S. Khemlani and Daniel M. Oppenheimer
Princeton University

Discounting is a phenomenon in causal reasoning in which the presence of one cause casts doubt on another. We provide a survey of the descriptive and formal models that attempt to explain the discounting process and summarize what current models do not account for and where room for improvement exists. We propose a levels-of-analysis framework organized around 2 types of models of causal discounting: computational and algorithmic models. Theories of causal discounting at the computational level attempt to provide normative, prescriptive explanations for discounting behavior, and they build on other normative frameworks like formal logic and probability theory. However, they tend not to focus on how those computations are carried out. Theories of discounting at the algorithmic level focus on the functions and representations from which discounting behavior emerges (i.e., they examine how problems are solved). We use this framework to identify gaps in the current literature and avenues for future model development.

Keywords: causal discounting, attribution, levels of analysis

Imagine you purchased a new automatic sprinkler system and you are anxious to see if it worked overnight. You walk outside and see that the grass is wet, and you assume the sprinklers must have gone off. However, when you read in the morning paper that it rained last night, you lose confidence that your sprinklers watered the lawn, and you think that the dampness may have been due to the rain instead. This kind of reasoning is known in the psychological literature as *causal discounting* (Einhorn & Hogarth 1986; Kelley, 1972b). The two potential causes of the wet lawn (the sprinklers and the rain) are not mutually exclusive—it is just as likely to rain whether or not the sprinklers functioned properly. Still, the presence of one cause casts doubt on the presence of the other. Recent research has shown that causal discounting plays a role in many cognitive processes. For example, discounting has been demonstrated in domains such as categorization judgments (Oppenheimer & Tenenbaum, 2010; Rehder, 2003), categorical induction (Oppenheimer & Frank, 2008), memory (Whittlesea & Williams, 1998, 2001a, 2001b), intelligence judgments (Oppenheimer, 2006), social judgments (Kelley, 1972b), and frequency estimations (Oppenheimer, 2004; Schwarz, 1998).

Despite several excellent reviews of the literature on discounting (McClure, 1998; P. H. Miller & Aloise, 1990; Morris & Larrick, 1995), to date there have been no comprehensive reviews of how different models of causal attribution explain discounting. The lack of an organizing structure has made it more difficult to identify gaps in the literature and redundancy among extant models. However, a natural organization emerges when analyzing the explanations for discounting. Models of discounting tend to draw inspiration from distinct *levels of analysis* (Marr, 1982).

Marr (1982) considered three levels at which to specify an information-processing system. The *computational level* specifies what is being computed and what problems are being solved, but it does not explicate the methods by which the problems are solved. The *algorithmic level* specifies the representations for the inputs and outputs of the system as well as how the inputs are transformed into the outputs. Finally, the *implementational level* specifies how the algorithms can be physically realized. These levels of analysis can serve as an organizing framework to inspire as well as constrain extensions of causal discounting research. Models of discounting tend to be specified at either the computational level or the algorithmic level, and we know of no implementational accounts of discounting. Thus, we focus our review on the computational and algorithmic levels of analysis. We then show the utility of the framework by examining how it brings to light novel ways in which researchers might explore discounting, in terms of modeling approaches as well as testing strategies to distinguish among various theories.

We begin by highlighting important distinctions among different types of attribution theories. Next, we briefly examine prominent empirical findings in the area. We then propose a framework

Sangeet S. Khemlani and Daniel M. Oppenheimer, Department of Psychology, Princeton University.

This article is based on work supported under National Science Foundation Graduate Research Fellowships to both authors. We thank Nicholas Davidenko, Tom Griffiths, Phil Johnson-Laird, Chip Heath, Ching Kao, Derek Koehler, David Mackenzie, Teenie Matlock, Benoit Monin, Michael Morris, Michael Ramscar, Itamar Simonson, Barbara Spellman, Josh Tenenbaum, Mike Timbol, Barbara Tversky, and Bruce Whittlesea, and the entire Oppenheimer Lab for their insightful suggestions, and we thank Vanessa Baker, Katherine Kixmiller, Peter Finlayson, David Mackenzie, and Dara Wathanapaisal for their assistance. Special thanks to Joanne Kane, Joe McGuire, and Abby Sussman for their careful critiques of an earlier version of this article. And, of course, thanks to mom and dad for everything.

Correspondence concerning this article should be addressed to Sangeet S. Khemlani, Green Hall, Department of Psychology, Princeton University, Princeton, NJ 08540. E-mail: khemlani@princeton.edu

for understanding the motivations and theoretical similarities between major models of causal attribution and apply the framework to generate new models of discounting. We conclude by discussing instances in which the current models fail to account for available data, proposing several novel approaches that might fill these gaps and describing challenges for future models of discounting.

What Is Discounting?

Theorists do not agree about what constitutes causal discounting. Therefore, providing a single comprehensive definition is challenging. On the broadest level, discounting can be defined as a situation in which the presence of one cause casts doubt on another. This can be operationalized in many ways, so it is important to clarify what we mean by discounting in the context of this review. We distinguish between the cause that is being discounted (the *original cause*) and the other causes that provide the impetus for discounting (the *alternative causes*).[1]

The causal reasoning literature contains two major areas of focus: how causal relationships are learned and how people make causal attributions. The first investigates how individuals induce whether an occurrence is a viable candidate as a cause for a particular event. It explains how one knows that wet grass can be caused by either rain or sprinklers. Psychologists investigating causal learning have examined how covariation between variables leads one to infer that an event can be produced by a particular cause (e.g., Cheng & Novick, 1992; Kelley, 1967; Novick, Fratianne, & Cheng, 1992; see also Hattori & Oaksford, 2007, for a comprehensive analysis of covariation detection). The word *discounting* has been used in the literature to refer to a decrease in the evaluation of the strength of the original cause after learning about the presence of an alternative cause (Goedert, Harsch, & Spellman, 2005; Kelley, 1972a; Laux, Goedert, & Markman, 2010). Most models of causal learning make no claims about predicting or explaining discounting (but see Goedert et al., 2005; van Overwalle, 1998) and thus are beyond the scope of this review. The second area of focus examines how, in a particular situation, people ascribe a cause to a particular event. That is, after observing a specific instance of an event, what causal attributions do individuals make? This approach explains how one determines which of those candidate causes led to the wet lawn this particular morning. In this article, we limit our discussion to causal attribution.

Another important distinction in the literature is between *discounting* and *explaining away*. Imagine that an individual has a belief in the likelihood of a given cause occurring, which we can call C_0. When that individual observes an event that is possibly an effect of that cause, most attribution theories argue that the individual's likelihood estimation for the cause will increase to C_1. In the presence of both an effect and an alternative cause, the individual's likelihood estimation for the original cause will drop. The question is whether it decreases back to C_0 or if it goes below baseline. If it decreases to C_0, the individual has explained away the original cause (Pearl, 2000; Wellman & Henrion, 1993); if it goes below C_0, it is known as discounting. There is empirical support for both explaining away (Shultz & Butowsky, 1977) and discounting (Kun, 1977; Shultz, Butkowsky, Pearce, & Shanfield, 1975; Smith, 1975), and the two concepts are often used interchangeably. Nevertheless, some of the models we reviewed can account only for explaining away (e.g., Kanouse, 1972; Kelley, 1972b; Pearl, 2000), and it is unclear whether explaining away and discounting are accomplished by the same cognitive mechanism. For a formal treatment of how explaining away can be construed as a special case of discounting, see Griffiths (2001).

Finally, we draw a distinction between discounting *likelihood* and discounting causal *strength*. In the former, the presence of alternative causes leads reasoners to lower their confidence that the original cause is present. In the latter, the presence of the original cause is not questioned, but the context forces a person to refocus or re-allocate strength from one link in a causal chain to another. Although Kelley (1972b) addressed this distinction, almost every model since has ignored it (but see Hilton & Erb, 1996). For the purpose of this review, we focus on the discounting of likelihood, because that has been the focus of the majority of the discounting literature.

How Common Is Discounting?

McClure (1998) reviewed the empirical literature on discounting in a seminal article. He argued that the common methodology of a single bipolar scale—for example, with *person* on one end and *situation* on the other (e.g., Elig & Frieze, 1979; F. D. Miller, Smith, & Uleman, 1981; Thibaut & Riecken, 1955)—leads to artificial validation of discounting, and he observed that many discounting effects disappear when this confound is removed (e.g., Lalljee, Watson, & White, 1982; Wimer & Kelley, 1982).

McClure (1998) further claimed that discounting is less common than typically assumed (see also Rosenfield & Stephan, 1977). Although his review focused on studies where people seemed to underdiscount (e.g., Schustack & Sternberg, 1981; Shaklee & Fischhoff, 1982) or not discount at all (e.g., Leddo, Abelson, & Gross, 1984; Zuckerman, Eghrari, & Lambrecht, 1986), he emphasized these results as a means of combating assumptions about the ubiquitous nature of discounting (e.g., Hansen & Hall, 1985). McClure ended with an important lesson: Discounting is not a universal tendency and is, instead, influenced by cognitive biases, the interaction of multiple causes, and associations between causes. Researchers should not assume discounting in every causal scenario, and models of discounting should be sensitive to the different semantic and logical relations between causes and effects.

However, others have argued that discounting is a general psychological phenomenon and that it can be observed in many specific situations in daily life. For instance, if you are aware that others may discriminate against you or your group for a particular reason, you can discount negative evaluations coming from those sources (Crocker & Major, 1989; Dion, 1975; Weiner, 1985a). This occurs frequently after authors read the reviews of a rejected

[1] The nature of a cause is complex and nuanced and has been debated by linguists, psychologists, and philosophers. For example, some theorists draw distinctions between causes and enabling or disabling conditions (Cheng & Novick, 1991; Cummins, 1995; Mackie, 1965; McGill, 1989), and others distinguish between communicated explanations and attributed causes (Wilson & Sperber, 2004). Although it is beyond the scope of this article to go into detail on these distinctions, we would like to direct the attention of interested readers to several excellent discussions of the topic (Hewstone, 1989; Hilton, 1990; Jaspars, 1983).

manuscript. The authors might think that Reviewer *X* obviously did not like their manuscript because the reviewer has proposed an alternative theory. In doing so, the authors protect their self-esteem by discounting one cause of a bad review (the quality of the article) in the presence of an alternative cause (the reviewer's supposed bias). Similarly, discounting can occur when individuals experience stereotyping, discrimination, and prejudice (Branscombe, Schmitt, & Harvey, 1999; Hewstone, 1994; Johnston & Hewstone, 1992; Schmitt & Branscombe, 2002; Schmitt, Branscombe, Kobrynowicz, & Owen, 2002). Discounting behavior can be observed early in development (Ali, Schlottmann, Shaw, Chater, & Oaksford, 2010; Fincham, 1983; P. H. Miller & Aloise, 1990; Sedlak & Kurtz, 1981). Moreover, researchers have developed alternative methods for ensuring that patterns consistent with discounting are not artifactually produced (e.g., Oppenheimer & Monin, 2009). Although discounting does not manifest in every causal context, it nonetheless appears to be a pervasive phenomenon.

Is Discounting Normative?

Although a great deal of research has demonstrated discounting under various conditions, there has been controversy about whether discounting is a normative behavior or a bias (Morris & Larrick, 1995). Although explaining away is generally construed as normative behavior, it is not clear whether discounting beyond baseline is optimal. Some researchers have argued that people tend to underdiscount (Jones, 1979; Quattrone, 1982); others have argued that people overdiscount (Nisbett & Ross, 1980). Still others have argued that an appropriate amount of discounting occurs (Olson, 1992). To clarify the debate, Morris and Larrick (1995) derived a series of normative principles to determine the extent to which individuals discount to the correct degree.

Morris and Larrick (1995) drew attention to the four parameters that seem to determine the extent to which discounting is normative: prior probability, sufficiency, independence, and number of causes. Prior probability is the base rate of a potential cause occurring. Sufficiency represents the probability that an effect will occur in the presence of the appropriate cause (e.g., although smoking causes cancer, there is not a perfect relationship between the variables; therefore, smoking is not a sufficient cause of cancer). Independence refers to the extent to which the alternative causes covary with the original cause. Finally, the number of alternative causes is the number of potential causes for a given effect.

Morris and Larrick (1995) did not intend their normative model to describe actual human behavior; they proposed that experiments and models of discounting could investigate whether their parameters constrain human discounting in the same way they constrain the extent to which discounting is normative. Many such models have been proposed, and in the next section, we present a framework that organizes the extant theories of causal discounting.

A Levels-of-Analysis Framework for Causal Discounting

Although researchers have explained discounting by appealing to a variety of cognitive processes, they have tended to draw inspiration from two distinct levels of analysis (Marr, 1982). First, researchers often specify their models at the computational level, the level of analysis that specifies what is being computed in a given process. In this sense, the computational level is related to rational analysis and the adaptive notion of human cognition (J. R. Anderson, 1990). Theories at the computational level attempt to provide normative, prescriptive explanations for discounting behavior, and they build on other normative frameworks like formal logic and probability theory. However, they tend not to focus on how those computations are carried out. Theories at the algorithmic level, the level of analysis that focuses on the functions and representations from which discounting behavior emerges, examine how problems are solved. They also examine related issues, such as what makes solutions efficient, the degradation and perturbation of performance, what makes some problems difficult and others easy, why some problems are solved quickly and others take a long time, and similar concerns. Researchers working at the algorithmic level often look for analogues of discounting in other cognitive processes. Such theories explore whether linguistic, heuristic, model-based, or other reasoning strategies map onto, or form the basis of, discounting behavior. These theories are often characterized by careful analyses of the representations and cognitive operations involved in causal discounting.

Our framework explicitly acknowledges the levels of analysis that serve as sources of inspiration for accounts of discounting. The framework allows researchers to identify and characterize novel models of discounting and to compare predictions across models. Formal models (i.e., those that are specified as effective procedures; Kleene, 1967) are privileged over more intuitive models because they can be realized by a series of computable steps and are thus more likely to yield specific, testable predictions. Indeed, for several models we examine, discounting behavior emerges as a result of the operations of a corresponding computer program. However, formal models are specified roughly in proportion to their generality: More specific models tend to describe fewer phenomena. It follows that some of the earliest, nonformal theories of discounting are also the most general. Therefore, we begin by examining the initial accounts of discounting and the insights they provided to later research.

Early Theories of Discounting

Initial theories of causal attribution and discounting described the phenomena without recourse to formalization. They served to identify and document discounting rather than to explain why it occurred. These early forays were useful for subsequent accounts of discounting because of the lucidity, intuitiveness, and clarity of description of qualitative phenomena.

Jones and Davis's (1965) *correspondent inference model* was the first observation of discounting as a phenomenon within causal reasoning. Jones and Davis focused on the attribution of behaviors and argued that inferences would be stronger for undesirable behaviors than for desirable behaviors, because far fewer alternative explanations would come to mind for undesirable behaviors. Students who write papers might do so for several reasons: They may like writing, have an impending deadline, or wants to get ahead at work. A student who pulls an all nighter to finish a paper usually does so because of an impending deadline, as the other explanations lack the strength to account for the undesirable evening. Correspondent inference theory is more useful for its historical perspective on the discovery of discounting than for any insight into the process, because it is primarily descriptive rather than explanatory in nature.

Another influential early approach to discounting is Kelley's (1972b) *causal schemata model*. Kelley hypothesized that individuals have impressions or schemas of how causes tend to interact to produce effects. Although Kelley never described the process of how these impressions were formed, he held that they were inferred from two types of experiences: observations of cause–effect relationships and direct manipulations of causal factors.

Kelley (1972b) described several types of causal schemata, the most relevant of which involves multiple sufficient causes.[2] According to Kelley, discounting occurs in this situation because, in the presence of an alternative cause, there is no need to reconsider the presence of the original cause because the effect is already accounted for. Thus, Kelley provided the original description of the notion of explaining away.

Models of Discounting at the Computational Level

The models of casual attribution proposed by Jones and Davis (1965) and Kelley (1972b) offer formative insights into causal attribution, how it might be studied, and how it manifests in daily life. Researchers have since attempted to find analogues of these ideas by specifying the functions being computed when discounting occurs. Such theories offer explanations at the computational level, and are characterized in several ways: First, the theories are often prescriptive and normative; that is, they explain why discounting ought to take place given certain constraints. Second, the theories tend not to specify the representations necessary for discounting behavior to be realized. They instead focus on the functions computed to achieve discounting. Morris and Larrick's (1995) seminal treatment falls under the category of a theory at the computational level, and in this section we review several others. Table 1 provides an overview of the models of causal discounting we examine in this section.

Discounting as Logical Deduction

Harkening back to Mill's (1843) notion of necessity and sufficiency, one prominent approach to causal attribution is the *natural logic model* (Hewstone & Jaspars, 1987; Jaspars, Hewstone, & Fincham, 1983). The natural logic model builds on Kelley's (1967) proposals and posits that individuals are equipped with a set of formal rules of inference that allow them to prove which possible causes are necessary or sufficient to account for a particular event. Information about consensus, distinctiveness, and consistency is coded as a set of conditions on which to base causal attributions. Although Jaspars et al. (1983) never directly discussed discounting, a natural extension of their theory can account for discounting through the principal of necessity. Consider the example of the wet lawn. The observer might think initially that the dampness is due to the sprinklers. Once the observer learns that it rained, then it is no longer necessary to attribute the dampness to the sprinklers—the lawn would be wet even in the absence of sprinklers. According to the natural logic model, it would therefore not be attributed to the sprinklers.

A similar model that is also based on logical calculations is Kruglanski's (1980) *lay epistemic model*. The model assumes that causal attribution is part of a grander scheme of knowledge acquisition. Accordingly, Kruglanski laid out a theory of knowledge acquisition, from which he derives a number of predictions about attribution. Kruglanski argued that acquiring knowledge should be thought of as selecting the valid proposition from a set of mutually exclusive propositions. According to the lay epistemic model, people prefer to use informative evidence, that is, evidence that rules out possibilities, when making attributions (Bar-Hillel & Carnap, 1953).

Kruglanski (1980) used informativeness to explain discounting by noting that a causal hypothesis loses strength in the presence of an alternative cause because the alternative cause provides evidence that rules out the original hypothesis. After all, the competing cause is consistent with itself (by identity), and because lay epistemic theory requires propositions to be mutually exclusive, the competing cause is inconsistent with the original hypothesis. As such, confidence in the original hypothesis decreases. Although Kruglanski discussed the importance of informativeness, the assumption of mutual exclusivity does most of the work in explaining discounting for lay epistemic theory.

Probabilistic Theories of Discounting

Probabilistic theories of discounting assign each cause a probability and examine the dynamics by which such probabilities shift when new information is learned. For instance, Morris and Larrick's (1995) subjective probability model of discounting uses Bayes's theorem to identify beliefs as subjective probabilities. Another broad probabilistic framework for attribution theory is Medcof's (1990) *probability expectancy attribution theory* (PEAT). The premise of PEAT is that expectations are central to the construction of attributions. Expectations are formalized through probability theory.

Medcof (1990) argued that base rates and conditional probabilities are stored in memory. For example, one might have a notion that the probability of the grass being wet at any given time is 20% and that the odds of the grass being wet after it has rained are 65%. In other words, $P(\text{wet grass}) = .20$, $P(\text{wet grass} \mid \text{rain}) = .65$. When a conditional probability is greater than a base rate, then there is a causal relationship; for example, if $P(\text{wet grass}) < P(\text{wet grass} \mid \text{rain})$, then the rain is a probable cause of wet grass. This assumption "is based upon lay people's tendency to use base perceptions of causality on predictability" (Medcof, 1990, p. 118).

According to Medcof (1990), when individuals attempt causal attributions after observing an event, they use one of four hypothesized modes of explanation: known characteristics, revised characteristics, new characteristics, and unknown cause. Prior understanding of causal mechanisms is often enough to explain the event. Therefore, when individuals have knowledge of the appropriate causal characteristics, they name one of the available causes. For revised and new characteristics, prior beliefs cannot account

[2] Kelley also discussed other causal schemata, such as the multiple necessary cause schema, but as Morris and Larrick (1995) observed, Kelley (1972b, p. 155) focused on the multiple sufficient cause schema and the compensatory schema as situations applicable to the discounting principle. Here, we focus on the multiple sufficient cause schema. For an analysis of how the compensatory schema applies to discounting, see Morris, Smith, and Turner (1998).

Table 1
Computational Models of Causal Discounting

Theory	Normative or descriptive	Parameters of discounting mechanism	Explaining away or discounting	Historical basis
Natural logic model (Jaspars et al., 1983)	Normative	Consensus, distinctiveness, consistency	Explaining away	Logic
Lay epistemic model (Kruglanski, 1980)	Normative	Exclusivity, informativeness	Discounting	Logic
Subjective probability model (Morris & Larrick, 1995)	Normative	Prior probability, sufficiency, independence, number of causes	Discounting	Probability theory
Probability expectancy attribution theory (Medcof, 1990)	Normative	Conservatism, parsimony, relative strength, number of causes	Discounting	Probability theory
Likelihood ratios (Ajzen & Fishbein, 1975)	Normative	Predictability, independence, presence and absence of causes	Explaining away	Probability theory
Bayesian networks (Pearl, 2000)	Normative	Graphical dependency relations	Explaining away	Probability theory

for the event. Therefore, people either revise current beliefs to explain the event or infer new characteristics about agents. In the case of unknown causes, people make no attribution or attribute the event to luck or chance.

Medcof (1990) proposed that individuals select an explanation mode on the basis of three principles: conservatism, parsimony, and relative strength. Conservatism is the belief that the world is orderly and predictable. Parsimony is the desire to make small modifications to current beliefs over large ones. Relative strength assumes that beliefs are held with different amounts of strength and that more strongly held beliefs are less likely to be modified than their weaker counterparts. As a natural consequence of these principles, the known characteristics mode of explanation is always the preferred mode. That is, if the situation can be explained by current knowledge, then the principle of conservatism is upheld. Parsimony is upheld by the fact that no modifications to beliefs are necessary, and because no beliefs have to be modified or abandoned, the principle of relative strength is irrelevant.

PEAT accounts for discounting through the proposed bias toward the known characteristics mode of explanation. When wet grass is observed in the absence of any obvious cause, a person cannot use the known characteristics mode of explanation to account for wetness (because the presence of a cause is not known). Instead, people must infer new or revised information about the situation—such as the sprinkler system—that moves their likelihood estimations of the sprinkler system above baseline. However, if wet grass is observed after reports of rain, observers are able to account for the wet grass with the strongest available facilitator, in this case the rain. Observers do not have to infer any new or revised characteristics about the wet grass to explain the information. Therefore, they simply use the known characteristics mode of explanation and the likelihood estimation of the sprinkler system remains at baseline.

Additionally, the algorithm that PEAT uses to calculate causal strength in the presence of multiple potential causes (e.g., when both the sprinkler system and the rain occur near the wet grass) is

$$ff(C) = ff(T)/n, \tag{1}$$

where $ff(C)$ is the amount of causal strength for cause C (Medcof, 1990, termed this *facilitory force*, hence the ff), n is the number of causes present, and $ff(T)$ is the total amount of causal strength necessary to create an event. This means that if there are multiple potential causes present when an event occurs, the event is attributed to each of them only partially. This is a second way in which PEAT explains discounting.

Equation 1 has several interesting properties that are worth mentioning. First, there is no way to incorporate the relative plausibility of alternative causes into the discounting process. The variable n represents the number of causes present, regardless of how reasonable it is to ascribe causation to those agents. This constraint seems at odds with our intuitions about discounting behavior. Imagine a situation in which a teacher is hit in the back of the head by a paper airplane. If the only person in the classroom is Nina Neutral, the teacher is likely to attribute the airplane attack to Nina, whereas if one of Nina's peers is also present, Nina might be discounted as a cause of the attack. However, it seems that different levels of discounting would occur if the peer is Nicholas Naughty than if the peer is Nancy Nice. PEAT's mechanism of discounting cannot distinguish between these cases. Further, Einhorn and Hogarth (1986) noted that ratio models cannot account for low causal ascription in the absence of alternative causes. In the end, both of these criticisms are empirical challenges; intuitive appeals are persuasive only to the extent that they inspire research to investigate potential shortcomings of the model.

Medcof's (1990) PEAT model is not the only model that attempts to capture the process of attribution through probability theory. Ajzen and Fishbein's (1975, 1978) approach uses Bayes's theorem and likelihood ratios to account for causal attribution. They began by stating Bayes's theorem with respect to two hypotheses: the presence of a causal factor (C) and the absence of that factor ($\sim C$):

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}, \tag{2}$$

$$P(\sim C|E) = \frac{P(E|\sim C)P(\sim C)}{P(E)}, \tag{3}$$

where E represents the event that might be attributed to cause C. By dividing Equation 2 by Equation 3, one is left with a particularly useful form of Bayes's theorem:

$$\frac{P(C|E)}{P(\sim C|E)} = \frac{P(E|C)}{P(E|\sim C)} \times \frac{P(C)}{P(\sim C)}. \tag{4}$$

In this equation, $P(E|C)/P(E|\sim C)$ is known as the likelihood ratio. This ratio represents how much the event favors one cause over the others. As the likelihood ratio increases, the odds are revised to favor a given causal hypothesis. Ajzen and Fishbein discussed in great detail how a variety of causal phenomena can be explained with this equation. Discounting is particularly easy to explain in this framework because the effects of alternative causes impact the likelihood ratio directly. Specifically, the presence of a plausible alternative cause increases $P(E|{\sim}C)$. Because this is the denominator of the likelihood ratio, the odds of the event being attributed to cause C are therefore decreased.

This approach has been criticized on several counts (Fischhoff & Lichtenstein, 1978). First, Fischhoff and Lichtenstein (1978) questioned the implicit normativity of using the likelihood ratio of Bayes's theorem to capture causal attribution. By using the likelihood ratio as an evaluation of causality, one captures the predictability of C from the occurrence of E, which does not necessarily imply causality, just as correlation does not necessarily imply causation (this criticism applies to PEAT as well). Second, Fischhoff and Lichtenstein criticized the descriptive adequacy of the model. They noted that if the potential causes are not independent, the complexity of the Bayesian calculations increases dramatically as the number of potential causes increases. Further, they described a tradition of research in heuristic judgments in which Bayesian models not only fail to predict human performance but fail to capture the important determinants of decision processes as well (for a review, see Slovic & Lichtenstein, 1971). However, as the next section shows, these critiques have not deterred theorists from proposing Bayesian accounts of causal reasoning.

Bayesian Networks

Bayesian networks (Charniak, 1991; Pearl, 2000) provide an alternative way of invoking Bayes's theorem to account for attribution. Bayesian networks are a popular way of modeling causal systems and constructing computational agents because they allow the agent to calculate the probabilities of all possible circumstances in the network from a much smaller subset of probabilities. Bayesian nets consist of nodes and causal arrows. The prior probabilities of the top level of the network are specified, and subsequent nodes in the network are described in terms of conditional probabilities. In other words, the system computes the odds of a child node occurring on the basis of the presence or absence of each cause in the preceding parent layer of nodes. For example, imagine a network in which sprinklers and rain cause wet grass, and wet grass in turn causes somebody's shoes to get wet. The top layer would consist of two nodes, *sprinkler* and *rain,* with the prior probability of each specified. The next layer would have a single node, *wet grass.* For this layer, four conditional probabilities would be specified: P(wet grass | sprinkler), P(wet grass | rain), P(wet grass | sprinkler, rain), and P(wet grass | ~sprinkler, ~rain). Finally, the bottom layer of this simple network would have a single node, *wet shoes.* The specification for this layer would require two pieces of information: P(wet shoes | wet grass) and P(wet shoes | ~wet grass).

Bayesian networks such as this one allow one to calculate the probability of any node on the network being true given the observed presence or absence of a piece of evidence (or any combination of evidence). For example, after observing a person with wet shoes, P(wet shoes) = 1, one could calculate the odds of wet grass, and through that calculation, the odds of it having rained. Bayesian networks often consist of thousands of nodes that can be re-evaluated each time new evidence comes in (Charniak, 1991).

Discounting is a natural consequence of a Bayesian network. If there are multiple potential causes leading to a single event, then the presence of that event leads to an increase in the probability of each of the potential causes being present. However, if one of the causes of the event has also been observed, then the nonobserved causes return to base probability (for a proof, see Pearl, 2000). Although Bayesian networks are a promising way of accounting for attribution and discounting, they too fall prey to Fischoff and Lichtenstein's (1978) critiques of descriptive adequacy, and they can account only for explaining away.

Summary of Computational Theories of Causal Discounting

Theories at the computational level specify the problems that discounting is meant to address. They stipulate the abstract inputs to a system that is designed to achieve a specific goal. The theories of discounting at the computational level do not explain how the mind computes the inputs or combines and manipulates them so that discounting is achieved. Instead, the selection of inputs constrains the kinds of strategies that might be used.

For instance, consider how Bayesian networks (Pearl, 2000) handle causal discounting. The theory considers discounting to be an emergent result of the calculations made by a network that is appropriately constructed. The theory states that for any network so constructed, explaining away (a special kind of discounting; see earlier discussion) is both appropriate and optimal. Bayesian network theory, like other accounts at the computational level, specifies the pieces of information needed to calculate a particular function (e.g., Bayes's theorem). The function describes what problem is being solved, and the pieces of information describe abstract constraints of an underlying cognitive mechanism. However, computational level theories do not specify how those pieces of information are represented or how they are integrated to execute the calculation of a given function. As a result, such theories cannot account for deviations from optimality. In other words, in situations that might be expected to produce discounting, theories at the computational level cannot explain why discounting might not occur. To address some of these issues, a number of models have been developed at the algorithmic level of analysis.

Algorithmic Accounts of Discounting

The models discussed in the previous section allow researchers to test whether human performance conforms to abstract idealizations. Some researchers have taken another approach: They have considered human performance in domains such as language comprehension, decision making, reasoning, and memory as a guide with which to explore the algorithmic processes involved in discounting behaviors. These theories are characterized by specifying the representations, as well as the transformations on those representations, necessary to produce discounting behavior. In this section, we review algorithmic models of discounting, and Table 2 provides an overview of these models.

Table 2
Algorithmic Models of Causal Discounting

Theory	Normative or descriptive	Parameters of discounting mechanism	Explaining away or discounting	Historical basis
Abnormal conditions focus model (Hilton & Slugoski, 1986)	Descriptive	Consensus, distinctiveness, consistency	Explaining away	Language
Satisficing model (Kanouse, 1972)	Descriptive	Satisficing, effort, parsimony	Explaining away	Language
Hydraulic model (Nisbett & Ross, 1980)	Descriptive	Causal strength resources, number of causes	Discounting	Game theory
Variant-effect model (Kun et al., 1980)	Descriptive	Causal strength resources, conservation, number of causes	Discounting	Physics
Dynamics model (Wolff, 2007)	Descriptive	Models	Explaining away	Physics
Counterfactual reasoning by proxy theory (Lipe, 1991)	Descriptive	Covariation, strength of alternative explanations	Discounting	Reasoning
Cues-to-causality model (Einhorn & Hogarth, 1986)	Descriptive	Temporal order, context, causal strength, covariation, similarity, attention	Discounting	Heuristics and biases
Mental models theory (Goldvarg & Johnson-Laird, 2001)	Descriptive	Semantics of causal relations, number of causes, preferred models	Discounting	Reasoning
Explanatory coherence by harmony optimization (Thagard, 1989)	Descriptive	Coherence, inhibitory links between nodes, number of causes	Discounting	Connectionism
Feedforward connectionist networks (van Overwalle, 1998)	Descriptive	Delta learning rule	Neither (blocking)	Connectionism

Linguistic Approaches to Causal Attribution

Hilton and Slugoski's (1986) *abnormal conditions focus model* argues that causal attribution is based on the same premises as conversational implicature. In particular, Hilton and Slugoski focused on the Gricean maxim of quantity, which states that speakers should not say things that listeners already know (Grice, 1975). Hilton and Slugoski extended this principle to attribution by arguing that attributers behave like speakers; they avoid mentioning causes that listeners already know. They offer an example of an individual who is asked to determine the cause of a particular pregnancy. The person would prefer to ascribe the pregnancy to a broken condom than to intercourse even though intercourse is necessary for pregnancy, because he assumes that the listener already knows that intercourse needs to occur. Therefore, by the maxim of quantity, the attribution should be made to the abnormal condition of a broken condom. However, Gricean principles can be overridden; they are descriptive rather than normative, and individuals may choose to violate them for specific purposes and in appropriate contexts.

An extended account of Hilton and Slugoski's (1986) use of Gricean principles can explain discounting as well. The maxim of quantity predicts that people should not give more information than required (Clark, 1996; Grice, 1975). Therefore, if speakers provide alternative causes, listeners will assume that the alternative causes are relevant to the discussion (otherwise they would not have been mentioned) and will explain away the original cause.

A lesser known model of discounting comes from Kanouse's (1972) treatise on language and attribution. Kanouse drew on linguistic data to argue for a satisficing mechanism and posited that individuals attempt to make causal inferences until a minimal criterion of explanation has been reached, at which point they stop making inferences. Kanouse contended that attributing a cause to an event is effortful and that people put forth cognitive effort to find a causal attribution when an event is completely unexplained but put much less effort into finding further and potentially better attributions once a minimally acceptable cause has been found. This satisficing hypothesis can also be extended to explaining away: Once a single cause has been identified, other causes appear less likely because an individual has already exceeded the threshold for attribution. Kanouse expressed this extension by stating his belief that unitary events are likely to be perceived as having unitary causes. Both the abnormal conditions focus model and Kanouse's *satisficing model* are limited to explaining away causes and cannot account for situations in which individuals discount the original cause.

Causal Resources and Physical Forces

Nisbett and Ross's (1980) *hydraulic model* of causation describes the attribution process in game theoretical terms. They claimed that individuals behave as though different plausible causal explanations for an event compete with each other in a zero–sum game. Discounting is handled by the hydraulic model by noting that if there are limited causal strength resources available, then (assuming the original cause is at ceiling) when an alternative cause with nonzero strength is added to the system, the original cause must lose strength so that the system does not exceed the total amount of causal force available. The model implies that one needs to represent a supply of causal resources for each event as well as the causal strength of each individual cause. One limitation of this approach is that Nisbett and Ross did not specify what happens when two or more alternative causes are added. Which causes lose strength? Is it the original cause, the alternative causes, or both? Depending on how the supply of causal resources is manipulated, the problem may become computationally intractable.

Kun, Murray, and Sredl (1980) provided a similar account of discounting; they drew an analogy to physics with the notion of conservation of causal force. Like Nisbett and Ross (1980), Kun et al. argued that there is a limited amount of total causal force available to explain a given event. However, their model is more developed in that it expands on Kelley's (1972b) notion of effect strength. According to Kun et al.'s *variant-effect model*, active causes generate causal forces toward the event. As the causal force increases, the strength of the effect increases as well. If multiple causes are present, each contributes force to the total and thus creates a stronger effect. For example, the presence of rain creates wet grass, but the presence of rain and the activated sprinkler system creates extremely soggy grass. Thus, Kun et al.'s model uses causal forces as its central representation and operations on those forces as a mechanism to explain causal attribution. Discounting follows from Kun et al.'s analysis in a clear-cut manner. If an effect is mild, then a single strong cause accounts for it, and no further force is necessary. All other causes are then discounted. Thus, discounting occurs if the effect appears to be due to a single cause or in situations where causal forces do not combine in an additive fashion (e.g., binary rather than continuous effects).

Wolff's (2007) *dynamics model* of causation holds that individuals represent causal assertions in a way that partially mimics the way causes and effects take place in the real world. The account is similar to that of Kun et al. (1980). However, it does not construe causation as the conservation, transfer, or exchange of causal strength. Instead, it construes causation as the interactions between vectors of causal forces. In particular, the account represents causation as a model in which a pattern of forces between an *affector* (a cause) and a *patient* yields an *end state* (the effects on the patient). Various causal relations like *cause*, *prevent*, *enable*, and *despite* emerge as a result of the dynamics between forces in the represented model. For example, suppose a toy boat is headed in a particular direction by the slow current of a river. A strong wind moving in the direction of the current can be said to *cause* the boat to move down the river. Similarly, a strong wind against the current can *prevent* the boat from moving down the river. A weak wind against the current might not perturb the boat from its course. Therefore, the boat moves down the river *despite* the force of the wind. All these situations describe interactions between affectors represented by force vectors. The wind vector acts on the current vector to bring about different end states depending on the relative force of each vector.

The dynamics model has not dealt with causal discounting explicitly, but the model can be extended to account for it: Initially, a force vector represents the original cause, but the presence of an alternative cause results in the same end state. Therefore, reasoners infer that the first cause is a vector of zero force. Recent advances of the dynamics model extend it to handle causal chains (Wolff, Barbey, & Hausknecht, 2010), which may shed light on additional mechanisms of discounting.

Counterfactual Reasoning

One way of looking at the problem of causal attributions is through the lens of counterfactual reasoning (i.e., the question of whether the sprinkler caused the grass to be wet can be reframed as whether the grass would have been wet had the sprinkler not gone off). Psychologists have long recognized the role of counterfactual reasoning in causal judgments (e.g., Kahneman & Miller, 1986; Wells & Gavanski, 1989), and more recently they have formalized it in mathematical terms (Lipe, 1991).

Lipe (1991) observed that answering counterfactual questions may be just as difficult as answering causal attribution questions; in actual life, an observer would rarely know the answers to counterfactual questions. To deal with this, Lipe created a model of *counterfactual reasoning by proxy*, in which she uses covariation data and alternative explanations to formalize how counterfactual causal reasoning might work. Lipe hypothesized a two by two contingency table in which a possible cause either does or does not occur and a given effect does or does not occur. She argued that one cell, in particular, is relevant to counterfactuals: when the event occurs without the hypothesized cause. Other cells are also relevant to the counterfactual question, but it is the relative size of the $\{E, \sim C\}$ cell that is central to counterfactual reasoning. Lipe suggested that the crucial comparison is between the $\{E, \sim C\}$ cell and the $\{E, C\}$ and $\{\sim E, \sim C\}$ cells. The relative strengths of these three cells are then calculated to assess the strength of a given hypothesis as a potential cause of an event.

Lipe (1991) took other alternatives into account, using the following formula:

$$L(C) = f[S(C) - kA], \tag{5}$$

where $L(C)$ is the judged likelihood that cause C was responsible for an event, $S(C)$ is the strength of evidence of the hypothesis that C caused the event, A is the strength of evidence that alternative causes were responsible for the event, k is a constant that serves as an adjustment factor, and f is a monotonically increasing function. The equation was designed to account for discounting and does so because the presence of alternative causes in the system would increase the value of A, which would decrease the overall likelihood of the original hypothesis.

Heuristics Approaches to Causal Attribution

Einhorn and Hogarth's (1986) *cues-to-causality* approach draws from findings in the judgment and decision making literature to model causal reasoning. Einhorn and Hogarth assumed that there are a variety of cues that indicate causality. These cues can be combined in a semicompensatory fashion to make causal attributions. The model discusses six cues that are central to attribution: covariation, temporal order, spatial contiguity, similarity of cause and effect, causal chain strength, and difference in background. Covariation is used in the statistical sense. Although correlation does not imply causation, the fact that people need to be frequently reminded of this truth suggests that covariation is often used as a cue for causation. People also believe that effects follow their causes in time. Therefore, temporal order plays a central role in causal attribution. Spatial contiguity is the notion that causes (especially physical causes) need to be proximal to the events they bring about. Similarity is related to the concept that there should be some sort of physical resemblance between a cause and effect. For example, physical similarities between parents and their children point to the parents as being causal agents of the child. The concept of causal chains is rather complex compared with the previous cues, but it boils down to the idea that the attributer must be able to imagine some way for a potential cause X to influence event Y in order to believe that X caused Y. Finally, difference in background is the notion that causes are more likely to be attributed to unusual things than things that are part of the normal background. Einhorn and Hogarth illustrated this concept with the example of a watch face breaking after being hit by a hammer. Because it is unusual for a watch to be hit by a hammer, causation is typically attributed to the hammer strike. However, if this hammer strike occurs in a watch factory during a test for defects, the hammer strike becomes part of the background, and the cause is now attributed to a problem in the glass face. Thus, an important cue to causation is how different an agent is from what is typically occurring in the background (i.e., how abnormal it is, a concept elaborated on by Hilton & Slugoski's, 1986, abnormal conditions focus model; see earlier).

Einhorn and Hogarth (1986) combined these cues in the following equation:

$$s(C, E) = Q_T Q_B Q_L(\lambda_C Q_C + \lambda_S Q_S). \tag{6}$$

In this equation $s(C, E)$ represents the extent to which C is considered to cause event E. Q_T is a binary term representing temporal order, with $Q_T = 1$ if C precedes E and $Q_T = 0$ if C follows E. Q_B is the extent to which C differs from the background on a $0 \leq Q_B \leq 1$ scale. Q_L is the causal chain strength as computed by the number of links in a chain necessary to connect C to E as a function of spatial contiguity. Q_C is a measure of covariation, and Q_S is a measure of similarity between cause and effect. Finally, the λs represent attentional weights. It is worth noting that spatial contiguity is not explicit in the equation but influences the attribution nonetheless through its impact on causal chains.

The equation is semicompensatory. Difference in background, causal chain strength, and temporal order are all noncompensatory in that if any of them is set to zero, then the causal strength of C is also zero. However, with the exception of temporal order, all Qs are continuous variables, and so weakness in one can be compensated for by strength in another. Although this semi-compensatory structure allows for a great deal of flexibility, the model in the form expressed earlier cannot account for discounting. Einhorn and Hogarth (1986) were aware of this and addressed discounting explicitly with a subtractive function:

$$S_k(C, E) = S_{k-1}(C, E) - w_k S(A_k, E), \tag{7}$$

where $S_k(C, E)$ is the strength of the cause C for E after k alternatives have been discounted, $S(A_k, E)$ is the strength of alternative A_k for E, and w_k is a constant that weights the A_kth alternative. As such, one evaluates each cause and subtracts away the strength of alternative causes, which leads to discounting. Einhorn and Hogarth's model is similar to Lipe's (1991) approach, except that Lipe's model uses covariation information that corresponds to counterfactual reasoning, whereas Einhorn and Hogarth's model uses a variety of heuristic cues combined in a semicompensatory fashion. The cues-to-causality model does not naturally imply discounting unless one explicitly adds a discounting heuristic. Such a heuristic may exist, but recent evidence suggests that discounting is an effortful process that goes away under cognitive load (Oppenheimer & Monin, 2009).

The cues-to-causality model differs from other theories of discounting in two ways. First, the mathematical terms in the model represent psychological factors, such as the difference in background and the similarity between cause and effect. As such, the model makes precise predictions of when humans discount on the basis of the specified subjective factors. For instance, the model predicts that a difference in background is proportional with the extent to which C is considered a cause of E. This prediction is testable if researchers can obtain subjective measurements of difference-in-background and causal strength.

Second, like Lipe's (1991) approach, the model posits that discounting is a subtractive process instead of a process that, for example, compares ratios or implements Bayes's theorem. Unlike ratio models, in which each successive alternative cause leads to smaller amounts of discounting, subtractive models make the testable prediction that additional alternative causes have similar levels of impact.

Mental Models Theory and Causality

Goldvarg and Johnson-Laird (2001) posited that when humans reason about causal relations, they construct *mental models*, or iconic representations of real, hypothetical, or imaginary situations (see also Hilton & Erb, 1996). They theorized that explanations are generated by building mental models of the causal relations of events and that the meaning of a causal relation between two separate events concerns what is possible in their co-occurrences. They accordingly distinguished among four causal relations: *C causes E*, *C prevents E*, *C allows E*, and *C allows not-E*. Consider the scenario in which the sprinkler system (*C*) caused wet grass (*E*). The theory suggests that humans construct the following mental models of possibilities consistent with the sprinkler system scenario:

sprinkler	wet grass
¬sprinkler	wet grass
¬sprinkler	¬wet grass.

The models are depicted as conjunctive sets of events and properties. Thus, the first model describes the scenario in which the sprinkler system went off and the grass is wet (i.e., the factual case). The other two models list counterfactual possibilities consistent with the fact that the sprinkler caused the grass to be wet.

The account also posits that individuals tend to hold only a single cause in working memory. This assumption predicts discounting behavior: Individuals represent original causes in memory but discard them in favor of alternative causes. That shift of focus then discounts the reasoner's belief in the presence of the original cause. In other words, discounting arises because of the limits of working memory capacity—only one model can be kept in mind at a time. Johnson-Laird observed that the theory does not systematically predict which model individuals tend to focus on, only that they focus on one model at a time (P. N. Johnson-Laird, personal communication). Nevertheless, in the domain of spatial reasoning, individuals show systematic biases toward representing a single, preferred mental model (Jahn, Knauff, & Johnson-Laird, 2007), and preferred models may carry over to a causal domain as well.

Constraint Satisfaction Models

Constraint satisfaction models (Holyoak & Thagard, 1989; Thagard, 1989, 1992) have also been applied to the problem of causal attribution and more specifically to discounting (L. C. Miller & Read, 1991; Read, 1987; Read & Marcus-Newhall, 1993). In particular, Thagard's (1989) *explanatory coherence by harmony optimization* (ECHO) model has shown promise in accounting for causal reasoning phenomena. The premise of the model is a network of loosely organized nodes that represent all of the potential concepts (e.g., causes and events) in a system. Each node is bidirectionally linked to every other node in the network. Activation spreads through the network along these links, and links can have an excitatory, inhibitory, or neutral weighting. Input information to one node then flows across links to related concepts. For example, after learning that the sprinklers had gone off, the *sprinkler* node in the network would be activated. The *sprinkler* node would lead to strong excitatory activation of the *wet grass* node, which in turn would have a strong reciprocal effect on the *sprinkler* node, a strong inhibitory effect on the *garden hose* node, and a weak inhibitory effect on the *water balloon* node (along with neutral effects on irrelevant nodes such as *purple*).

When a new input is introduced into the system, activation propagates through all the links in parallel, creating a new set of activation levels. On every iteration of this algorithm, each node's activation is calculated by combining the node's previous activation and the weighted combination of the activation of nodes to which it is linked. This process repeats itself through a constraint satisfaction process until, after many iterations, it converges on a set of activations that approaches the ideal compromise among the constraints that have been set by excitatory and inhibitory links.

The constraint satisfaction model is able to account for discounting through the use of inhibitory links between potential causes. That is, although the *sprinkler* node and the *rain* node are both positively linked to the *wet grass* node, they would be negatively linked to each other. Thus, if just *wet grass* were given as input, both *sprinkler* and *rain* would see increases in activation and be judged as likely causes. However, if both *wet grass* and *rain* were provided as input, the increase in activation that *sprinkler* would receive from *wet grass* would be countered by the inhibitory link with *rain*.

In a seminal article on formal accounts of discounting, Read and Marcus-Newhall (1993) adapted the ECHO software and compared the simulation with human data. They found that the model did an excellent job of accounting for a variety of causal reasoning effects, including discounting. They concluded that constraint satisfaction models are a useful abstraction for helping researchers predict when discounting will occur because the model "allows one to explicitly represent important aspects of the nature of relations among causes" (Read & Marcus-Newhall, 1993, p. 444).

One major shortcoming of Read and Marcus-Newhall's (1993) ECHO-based models of discounting is that they rely on inhibitory links between possible causes. According to the constraint satisfaction model, discounting occurs only if alternative causes are independent or negatively related. This prediction is testable, and although it has not been fully resolved, preliminary data suggest that negative associations between causes are not necessary for discounting to occur (Oppenheimer & Tenenbaum, 2010).

Constraint satisfaction can account for discounting in another way. If total activation is normalized, then increases in activation in one node could lead to decreases in other nodes because all activation is relative. (This is akin to Nisbett & Ross's, 1980, hydraulic model described earlier). However, this normalization approach suffers from the fact that adding multiple causes for a given outcome can actually lower the belief that the outcome occurred. Thus, if you are told that the sprinklers had gone off, it had rained, a pipe had burst, and somebody threw water balloons nearby (all of which have activations at ceiling), then there is not much activation left for the wet grass!

Feedforward Connectionist Networks

Alternative connectionist models to constraint satisfaction networks are *feedforward neural nets* (McClelland & Rumelhart, 1988; van Overwalle, 1998). These networks differ from constraint satisfaction models in that they have multiple layers, and links between nodes are unidirectional. Recently, van Overwalle (1998) applied these models to the problem of causal reasoning and discounting. As with constraint satisfaction models, each potential cause and effect is represented by a node. However, unlike the ECHO model, the network has multiple layers; nodes in the top layer represent events, whereas nodes in the bottom layer represent possible causes. Further, the feedforward network has nodes to represent conjunctions of possible causes. In these networks, activation flows in only one direction: from the cause nodes to the event nodes.

Unlike the ECHO model, which requires those weights to be set a priori by the experimenter (but see Read & Montoya, 1999), van Overwalle's (1998) model has built in learning algorithms to dynamically develop the link weights. The feedforward network uses a delta learning rule to acquire link weights. This rule is similar to the Rescorla–Wagner learning rule in animal conditioning (Rescorla & Wagner, 1972). In each iteration of learning, the activations of the event nodes are calculated with a weighted sum of the causes present in a given trial. This activation is then compared with the actual presence of the event (with a score of 1 if the event actually occurred and a score of 0 if it did not occur). The discrepancy between the predicted occurrence and the actual occurrence of the event is used to adjust the weights on the links between potential causes and events. Through numerous iterations of this process, the network eventually approximates ideal link weights for the causal environment. Once the ideal weights are calculated, one can examine causal attribution by activating each cause individually and determining the activation output of the given event. If a particular cause node leads to a high activation of the event node, then it is considered to be a likely cause.

van Overwalle (1998; van Overwalle & van Rooy, 2001) argued that feedforward networks explain discounting during the learning stage. When multiple causes are active in the presence of an event, they have to compete for the available adjustments of link weights. However, when there is only one cause active, there is no such competition, and the nonpresent cause is given a lower weight relative to the present cause. Such an explanation of discounting is problematic because the phenomenon that van Overwalle's model predicts is not, in fact, discounting. van Overwalle's models predict the learning of causal strengths; learning that one cause on its own leads to a given effect makes it harder to subsequently learn that other causes also lead to that effect. Although researchers have argued that this is equivalent to discounting (see Baker, Mercier, Vallee-Tourangeau, Frank, & Pan, 1993; Shanks, 1991), there does seem to be a distinction—namely, van Overwalle's model has no way of accounting for discounting once the information has already been learned (Read & Montoya, 1999). If a person has already learned that sprinklers and rain are legitimate causes for wet grass, then van Overwalle's model cannot account for how the knowledge that it had rained might decrease that person's confidence in the sprinkler's functioning to explain a particular instance of wet grass.

Summary of Algorithmic Theories of Causal Discounting

Algorithmic theories of discounting tend to specify the processes involved when individuals discount causes. Therefore, they often focus on the representations that are necessary to compute various functions. They acknowledge that certain processes are vulnerable to perturbation and distortion. Thus, they tend to be descriptive and not normative. Unlike theories at the computational level, algorithmic theories of causal discounting do not account for when and why discounting should occur. They instead specify how discounting occurs as a result of a series of computations.

Theories at the algorithmic level are often designed to specify more general causal mechanisms, such as the interpretation of causal assertions (Goldvarg & Johnson-Laird, 2001; Wolff, 2007). As a result, they can make explicit an underlying representational format of the causes being discounted. Not all theories at the algorithmic level explicitly specify representational formats, however. Therefore, one challenge for researchers is to identify which representations are compatible with the algorithmic processes specified by theories outlined in this section.

Using the Framework

We have organized existing models of causal reasoning around a levels-of-analysis framework on the basis of whether the models are specified at the computational or algorithmic level. By scrutinizing the framework for gaps, we can identify promising approaches to explaining discounting that have not, to date, been explored. In the present section, we identify two potential explanations of discounting: one computational theory and one algorithmic theory. Although the models we explore have not been construed as explanations of discounting before, by couching them in such a way we attempt to show how our framework can be used not just to catalogue known accounts of discounting but to generate new ideas and connections to other theories as well.

Exploring Computational Accounts: Quantum Probability

Researchers have explored classic probability theory in an attempt to account for discounting. Recently, some researchers have begun exploring how quantum probability might explain cognitive phenomena (Pothos & Busemeyer, 2009). Quantum probability is similar to standard probability theory, with a primary difference being that the distributive axiom is relaxed. In mathematical terms, that means

$$P(E \wedge (C \vee A)) =/= P((E \wedge C) \vee (E \wedge A)) \tag{8}$$

According to quantum probability, x and $\sim x$ can co-occur until one of these states is actually observed. Although this seems counterintuitive, a number of empirical results in physics suggest that this may be the way the world functions (Feynman, Leighton, & Sands, 1964). Quantum probability theory asserts that when neither potential cause has been observed, the reasoner is in a superimposed state where two mutually exclusive propositions can both be true. However, once one of these propositions has been observed, the state collapses to be consistent with the observation; measurement itself changes the nature of the probabilities. Thus, after learning that an alternative cause exists, the superimposed state that allowed for a high belief in the initial cause collapses and lowers the belief in the initial cause—a classic discounting effect.

Thus, causal attribution seems an area ripe for models of quantum cognition. By applying a rigorous treatment of quantum mathematics, researchers might unearth a number of interesting predictions and moderating factors.

Exploring Algorithmic Accounts: Memory and Discounting

Although researchers have sought inspiration from basic cognitive processes (e.g., language, counterfactual reasoning, decision making) to explain discounting, they have overlooked the possibility that discounting could arise from properties of memory. In particular, query theory (Weber et al., 2007) might be a useful approach (see also M. C. Anderson & Spellman, 1995). Query theory holds that when people make judgments, they consult their memory for relevant evidence and cues. The retrieval and processing of retrieved information occurs serially, and memories interfere with each other, such that retrieving one piece of information temporarily makes subsequent retrieval more difficult (M. C. Anderson & Bjork, 1994; Johnson, Häubl, & Keinan, 2007; Nickerson, 1984). Query theory thus implies that search order influences the effectiveness of evidence retrieval for various options.

Query theory can account for discounting in that the search for evidence is likely to be biased toward the candidate causes that are known to be present. Thus, in the presence of an alternative cause *A,* people retrieve information about how the event could be caused by *A*. They then retrieve information about how the event could be due to the original cause (*C*), but because of memory blocking, fewer pieces of evidence come to mind. Because less evidence supports the interpretation that *C* was responsible for the observed effect, people tend believe it less likely that *C* was responsible for the event. The model of discounting makes the novel prediction that discounting should go away if the search order is changed, an effect that has been shown for intertemporal choice (Weber et al., 2007) and the endowment effect (Johnson et al., 2007). This model also suggests that discounting should increase under time pressure, because evidence for *C* is sought after evidence for the alternative.

Future Directions in Causal Discounting

The causal attribution literature germinated from the areas of behavioral attribution (Heider, 1958; Kelley, 1972b) and attitudes (Bem, 1967; Jones & Harris, 1967). More recently, the study of causal ascription has been broadened to encompass the understanding of a range of causes in the real world (Einhorn & Hogarth, 1986; Pearl, 2000). The models described in this article represent a wide range of approaches to understanding causal attribution and discounting. We presented both descriptive and formal models that represent many different disciplines, including social psychology (e.g., Kelley, 1972b), developmental psychology (e.g., Kun et al., 1980), and cognitive psychology (e.g., Einhorn & Hogarth, 1986). Among the formal models, we have reviewed a diverse set of attempts to account for discounting, including subtraction models, ratio models, and connectionist accounts. One obvious future direction is to determine which of these models can best account for human behavior. In our tables, we helped parse out what the models require as input, and what they provide as output. Thus, the framework can serve as a guide for researchers who wish to discriminate among these theories.

Another major area of future investigation—which is equally pressing but has thus far been ignored—is determining the mechanism for causal search. That is, when and how do individuals retrieve potential causes from memory for use in causal ascription? All of the models presented in this article assume a clear set of alternative causes on which calculations can be made. In the real world, however, there are an inexhaustible number of events and an equally inexhaustible number of causes; when do we search for causal attributions, and what memory process drives these searches? Few studies have examined the factors that lead to spontaneous causal search. Weiner (1985b) reviewed the entire literature on spontaneous causal thinking and found only 17 articles that investigated this area. Only a handful of articles have examined the topic since Weiner's review. Weiner concluded that causal search is elicited by an unexpected event, a failure to attain a goal, or being explicitly asked for a cause but also acknowledges that that other attributes of a situation might lead to search as well.

The following question remains: How does search occur? The most developed models of causal search make few attempts to explain this issue. For example, Hastie (1984) described a seven-stage model of causal search, but the actual search through memory is described only as *information seeking* with no further elaboration. Shaklee and Fischhoff (1982) described the search through memory for information about causes, but participants were provided with a list of all possible causes in advance; the search for which causes might be relevant was never examined.

Perhaps the most promising approach to causal search comes from the *hypothesis generation model* (HyGene; Thomas, Dougherty, Sprenger, & Harbison, 2008). This model proposes that causal search is restricted to a small subset of causes previously experienced by the reasoner and suggested by observations. The reasoner holds and evaluates the causes in working memory, and they, in turn, guide the search for novel causes. Thomas et al. (2008) recognized that searching for causes may be a result of several interacting heuristics, and they are investigating whether systematic biases occur when searching for causes.

None of the models of causal reasoning described in the present review have any memory search component except for query theory. The subtractive and ratio models never describe what items will be considered as possible hypotheses, and although proponents of constraint satisfaction models (Read & Marcus-Newhall, 1993) and Bayesian networks (Pearl, 2000) may claim to account for causal search by having every possible cause available in the network, the computational requirements of such systems would be overwhelming.

The future of research into models of causal discounting should test the predictions of the models we have described and should incorporate or propose a memory search mechanism if progress in the modeling of causal attribution and discounting is to be made. Causal discounting research also appears to rest in understanding what role causal knowledge plays in a variety of other cognitive phenomena. For example, it has long been argued that causation plays a role in categorization (Murphy & Medin, 1985), and recent research has investigated the possibility that discounting occurs in this domain (Oppenheimer & Frank, 2008; Oppenheimer & Tenenbaum, 2010; Rehder, 2003). Other work has explored connections between causal reasoning and decision making (Hagmayer & Sloman, 2009), memory (Dennis, Lee, & Kinnell, 2008; Hemmer & Steyvers, 2009), language (Goldwater, Griffiths, & Johnson, 2009; Griffiths & Kalish, 2007; Wagenmakers et al., 2004), perception (Kersten, Mamassian, & Yuille, 2004; Schooler, Shiffrin, & Raaijmakers, 2001), metacognition (Alter & Oppenheimer, 2009; Oppenheimer, 2004; Schwarz et al., 1991), emotion (Schwarz & Clore, 1983), the consciousness of will (Wegner, 2004), and priming (Bargh, 1996; Erb, Bioy, & Hilton, 2002).

Despite these advances, few of the models we have reviewed can parry McClure's (1998) original criticisms of the ubiquity of discounting. Theories of discounting are incomplete if they cannot also predict when discounting will not occur. Future research should examine the semantic relations, pragmatic factors, and knowledge constraints that modulate or prevent discounting behavior from happening.

Conclusions

Causal reasoning is ubiquitous and fundamental to human cognition. In this article, we attempted to elucidate the process of causal attribution by examining one of its associated phenomena: causal discounting. We have reviewed the major theories and models that claim to account for discounting and have characterized the types of information each model accounts for and provides. We have distinguished computational-level theories from algorithmic-level theories, and we have provided some suggestions for how the field might use these distinctions to evaluate models of discounting.

Although some researchers have described the 1970s as the heyday of attribution research (Weiner, 1985b), there is still a great deal that researchers do not know. Advances in modeling techniques and a renewed emphasis on causal inference and attribution as an underlying principle for a host of cognitive processes suggest that discounting research may be entering a second renaissance. With the abundance of models available for consideration, researchers must recognize when one model casts doubt on another.

References

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin, 82,* 261–277. doi:10.1037/h0076477

Ajzen, I., & Fishbein, M. (1978). Use and misuse of Bayes's theorem in causal attribution: Don't attribute it to Ajzen and Fishbein either. *Psychological Bulletin, 85,* 244–246. doi:10.1037/0033-2909.85.2.244

Ali, N., Schlottmann, A., Shaw, A., Chater, N., & Oaksford, M. (2010). Causal discounting and conditional reasoning in children. In N. Chater & M. Oaksford (Eds.), *Cognition and conditionals* (pp. 117–134). Oxford, United Kingdom: Oxford University Press.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13,* 219–235. doi:10.1177/1088868309341564

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

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265–326). San Diego, CA: Academic Press.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100. doi:10.1037/0033-295X.102.1.68

Baker, A. G., Mercier, P., Vallee-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 414–432. doi:10.1037/0278-7393.19.2.414

Bargh, J. A. (1996). Automaticity in social psychology. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169–183). New York, NY: Guilford Press.

Bar-Hillel, Y., & Carnap, R. (1953). Semantic information. *British Journal for the Philosophy of Science, 4,* 147–157. doi:10.1093/bjps/IV.14.147

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74,* 183–200. doi:10.1037/h0024835

Branscombe, N. R., Schmitt, M. T., & Harvey, R. D. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology, 77,* 135–149. doi:10.1037/0022-3514.77.1.135

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine, 12,* 50–63. Retrieved from http://www.aaai.org/Magazine/magazine.php

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40,* 83–120. doi:10.1016/0010-0277(91)90047-8

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99,* 365–382. doi:10.1037/0033-295X.99.2.365

Clark, H. H. (1996). *Using language.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511620539

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96,* 608–630. doi:10.1037/0033-295X.96.4.608

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition, 23,* 646–658. Retrieved from http://mc.psychonomic-journals.org/

Dennis, S., Lee, M. D., & Kinnell, A. (2008). A Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language, 59,* 361–376. doi:10.1016/j.jml.2008.06.007

Dion, K. L. (1975). Women's reactions to discrimination from members of the same or opposite sex. *Journal of Research in Personality, 9,* 294–306. doi:10.1016/0092-6566(75)90004-5

Einhorn, H., & Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin, 99,* 3–19. doi:10.1037/0033-2909.99.1.3

Elig, T. W., & Frieze, I. H. (1979). Measuring causal attributions for success and failure. *Journal of Personality and Social Psychology, 18,* 105–115. doi:10.1037/0022-3514.37.4.621

Erb, H., Bioy, A., & Hilton, D. (2002). Choice preferences without inferences: Subconscious priming of risk attitudes. *Journal of Behavioral Decision Making, 15,* 251–262. doi:10.1002/bdm.416

Feynman, R., Leighton, R., & Sands, M. (1964). *The Feynman lectures on physics.* Reading, MA: Addison-Wesley.

Fincham, F. D. (1983). Developmental dimensions of attribution theory. In J. Jaspars, F. D. Fincham, & M. Hewstone (Eds.), *Attribution theory and research: Conceptual, developmental, and social dimensions* (pp. 117–164). London, United Kingdom: Academic Press.

Fischhoff, B., & Lichtenstein, S. (1978). Don't attribute this to Reverend Bayes. *Psychological Bulletin, 85,* 239–243. doi:10.1037/0033-2909.85.2.239

Goedert, K. M., Harsch, J., & Spellman, B. A. (2005). Discounting and conditionalization: Dissociable cognitive processes in human causal inference. *Psychological Science, 16,* 590–595. doi:10.1111/j.1467-9280.2005.01580.x

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25,* 565–610. doi:10.1207/s15516709cog2504_3

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition, 112,* 21–54. doi:10.1016/j.cognition.2009.03.008

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol 3. Speech acts* (pp. 41–58). New York, NY: Academic Press.

Griffiths, T. L. (2001). *Explaining away and the discounting principle: Generalising a normative theory of attribution.* Unpublished manuscript.

Griffiths, T. L., & Kalish, M. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science, 31,* 441–480. Retrieved from http://www.cognitivesciencesociety.org/journal_csj.html

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

Hansen, R. D., & Hall, C. A. (1985). Discounting and augmenting facilitative and inhibitory forces: The winner takes almost all. *Journal of Personality and Social Psychology, 49,* 1482–1493. doi:10.1037/0022-3514.49.6.1482

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology, 46,* 44–56. doi:10.1037/0022-3514.46.1.44

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

Heider, F. (1958). *The psychology of interpersonal relations.* New York, NY: Wiley. doi:10.1037/10628-000

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science, 1,* 189–202. doi:10.1111/j.1756-8765.2008.01010.x

Hewstone, M. (1989). *Causal attribution: From cognitive processes to collective beliefs.* Oxford, United Kingdom: Blackwell.

Hewstone, M. (1994). Revision and change of stereotypic beliefs: In search of the elusive subtyping model. *European Review of Social Psychology, 5,* 69–109. doi:10.1080/14792779543000020

Hewstone, M., & Jaspars, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology, 53,* 663–672. doi:10.1037/0022-3514.53.4.663

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107,* 65–81. doi:10.1037/0033-2909.107.1.65

Hilton, D. J., & Erb, H. (1996). Mental models and causal explanation: Judgments of probably cause and explanatory relevance. *Thinking and Reasoning, 2,* 273–308. doi:10.1080/135467896394447

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review, 93,* 75–88. doi:10.1037/0033-295X.93.1.75

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355. Retrieved from http://www.cognitivesciencesociety.org/journal_csj.html doi:10.1207/s15516709cog1303_1

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition, 35,* 2075–2087. Retrieved from http://mc.psychonomic-journals.org/

Jaspars, J. M. F. (1983). The process of attribution in commonsense. In M. R. C. Hewstone (Ed.), *Attribution theory: Social and functional extensions* (pp. 335–358). Oxford, United Kingdom: Blackwell.

Jaspars, J. M. F., Hewstone, M. R. C., & Fincham, F. D. (1983). Attribution theory and research: The state of the art. In J. M. F. Jaspars, F. D. Fincham, & M. R. C. Hewstone (Eds.), *Attribution theory: Conceptual developmental and social dimensions* (pp. 1–36). San Diego, CA: Academic Press.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 33, 461–474. doi:10.1037/0278-7393.33.3.461

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: 3. Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28,* 360–386. doi:10.1016/0022-1031(92)90051-K

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist, 34,* 107–117. doi:10.1037/0003-066X.34.2.107

Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 219–266). New York, NY: Academic Press.

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

Kahneman, D., & Miller, K. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153. doi:10.1037/0033-295X.93.2.136

Kanouse, D. E. (1972). Language, labeling, and attribution. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (pp. 192–238). Lincoln, NE: University of Nebraska Press.

Kelley, H. H. (1972a). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.

Kelley, H. H. (1972b). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271–304. doi:10.1146/annurev.psych.55.090902.142005

Kleene, S. C. (1967). *Mathematical logic.* New York, NY: Wiley.

Kruglanski, A. W. (1980). Lay epistemologic process and contents: Another look at attribution theory. *Psychological Review, 87,* 70–87. doi:10.1037/0033-295X.87.1.70

Kun, A. (1977). Development of the magnitude–covariation and compensation schemata in ability and effort attributions of performance. *Child Development, 48,* 862–873. doi:10.2307/1128335

Kun, A., Murray, J., & Sredl, T. (1980). Misuses of the multiple sufficient causal scheme as a model of naïve attributions: A case of mistaken identity. *Developmental Psychology, 16,* 13–22. doi:10.1037/0012-1649.16.1.13

Lalljee, M., Watson, M., & White, P. (1982). Explanations, attributions and the social context of unexpected behavior. *European Journal of Social Psychology, 12,* 17–29. doi:10.1002/ejsp.2420120102

Laux, J. P., Goedert, K., & Markman, A. (2010). Causal discounting in the presence of a stronger cue is due to bias. *Psychonomic Bulletin & Review, 17,* 213–218. doi:10.3758/PBR.17.2.213

Leddo, J., Abelson, R. P., & Gross, P. H. (1984). Conjunctive explanations: When two reasons are better than one. *Journal of Personality and Social Psychology, 47,* 933–943. doi:10.1037/0022-3514.47.5.933

Lipe, M. G. (1991). Counterfactual reasoning as a framework for attribution theories. *Psychological Bulletin, 109,* 456–471. doi:10.1037/0033-2909.109.3.456

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly, 2,* 245–264. Retrieved from http://www.press.uillinois.edu/journals/apq.html

Marr, D. (1982). *Vision: A computational approach.* San Francisco, CA: Freeman.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs and exercises.* Cambridge, MA: Bradford Books.

McClure, J. (1998). Discounting of causes of behavior: Are two reasons better than one? *Journal of Personality and Social Psychology, 74,* 7–20. doi:10.1037/0022-3514.74.1.7

McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology, 57,* 189–200. doi:10.1037/0022-3514.57.2.189

Medcof, J. W. (1990). PEAT: An integrative model of attribution processes. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 111–209). San Diego, CA: Academic Press.

Mill, J. S. (1843). *A system of logic.* London, United Kingdom: Longmans.

Miller, F. D., Smith, E. R., & Uleman, J. (1981). Measurement and interpretation of situational and dispositional attributions. *Journal of Experimental Social Psychology, 17,* 80–95. doi:10.1016/0022-1031(81)90008-1

Miller, L. C., & Read, S. J. (1991). On the coherence of mental models in persons and relationships: A knowledge structure approach. In F. Fincham & G. J. O. Fletcher (Eds.), *Cognition in close relationships* (pp. 69–99). Hillsdale, NJ: Erlbaum.

Miller, P. H., & Aloise, P. A. (1990). Discounting in children: The role of social knowledge. *Developmental Review, 10,* 266–298. doi:10.1016/0273-2297(90)90013-T

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

Morris, M. W., Smith, E. E., & Turner, K. (1998). Parsimony in intuitive explanations for behavior: Reconciling the discounting principle and preference for conjunctive explanations. *Basic and Applied Social Psychology, 20,* 71–85. doi:10.1207/s15324834basp2001_7

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316. doi:10.1037/0033-295X.92.3.289

Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition, 12,* 531–552. Retrieved from http://mc.psychonomic-journals.org/

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Novick, L. R., Fratianne, A., & Cheng, P. W. (1992). Knowledge-based assumptions in causal attribution. *Social Cognition, 10,* 299–333. Retrieved from http://www.guilford.com/pr/jnco.htm

Olson, J. M. (1992). Self-perception of humor: Evidence for discounting and augmentation effects. *Journal of Personality and Social Psychology, 62,* 369–377. doi:10.1037/0022-3514.62.3.369

Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science, 15,* 100–105. doi:10.1111/j.0963-7214.2004.01502005.x

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20,* 139–156. doi:10.1002/acp.1178

Oppenheimer, D. M., & Frank, M. F. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition, 106,* 1178–1194. doi:10.1016/j.cognition.2007.05.010

Oppenheimer, D. M., & Monin, B. (2009). Investigations in spontaneous discounting. *Memory & Cognition, 37,* 608–614. doi:10.3758/MC.37.5.608

Oppenheimer, D. M., & Tenenbaum, J. B. (2010). *Categorization as causal explanation.* Manuscript submitted for publication.

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

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of “rational” decision theory. *Proceedings of* *the Royal Society B: Biological Sciences, 276,* 2171–2178. doi:10.1098/rspb.2009.0121

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology, 42,* 593–607. doi:10.1037/0022-3514.42.4.593

Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology, 52,* 288–302. doi:10.1037/0022-3514.52.2.288

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65,* 429–447. doi:10.1037/0022-3514.65.3.429

Read, S. J., & Montoya, J. A. (1999). An autoassociative model of causal reasoning and causal learning: Reply to van Overwalle's (1988) critique of Read and Marcus Newhall (1993). *Journal of Personality and Social Psychology, 76,* 728–742. doi:10.1037/0022-3514.76.5.728

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27,* 709–748. doi:10.1207/s15516709cog2705_2

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Rosenfield, D., & Stephan, W. G. (1977). When discounting fails: An unexpected finding. *Memory & Cognition, 5,* 97–102. Retrieved from http://mc.psychonomic-journals.org/

Schmitt, M. T., & Branscombe, N. R. (2002). The meaning and consequences of perceived discrimination in disadvantaged and privileged social groups. *European Review of Social Psychology, 12,* 167–199. doi:10.1080/14792772143000058

Schmitt, M. T., Branscombe, N. R., Kobrynowicz, D., & Owen, S. (2002). Perceiving discrimination against one's gender group has different implications for well-being in men and women. *Personality and Social Psychology Bulletin, 28,* 197–210. doi:10.1177/0146167202282006

Schooler, L. J., Shiffrin, R. M., & Raaijmakers, J. G. (2001). A Bayesian model for implicit effects in perceptual identification. *Psychological Review, 108,* 257–272. doi:10.1037/0033-295X.108.1.257

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120. doi:10.1037/0096-3445.110.1.101

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review, 2,* 87–99. doi:10.1207/s15327957pspr0202_2

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61,* 195–202. doi:10.1037/0022-3514.61.2.195

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523. doi:10.1037/0022-3514.45.3.513

Sedlak, A. J., & Kurtz, S. T. (1981). A review of children's use of causal inference principles. *Child Development, 52,* 759–784. doi:10.2307/1129077

Shaklee, H., & Fischhoff, B. (1982). Strategies of information search in causal analysis, *Memory & Cognition, 10,* 520–530. Retrieved from http://mc.psychonomic-journals.org/

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 433–443. doi:10.1037/0278-7393.17.3.433

Shultz, T. R., & Butkowsky, I. (1977). Young children's use of the schemes for multiple sufficient causes in the attribution of real and hypothetical behavior. *Child Development, 48,* 464–469. doi:10.2307/1128640

Shultz, T. R., Butkowsky, I., Pearce, J. W., & Shanfield, H. (1975). Development of schemes for the attribution of multiple psychological causes. *Developmental Psychology, 11,* 502–510. doi:10.1037/h0076672

Slovic, P., & Lichtenstein, S. C. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744. doi:10.1016/0030-5073(71)90033-X

Smith, M. C. (1975). Children's use of the multiple sufficient cause schema in social perception. *Journal of Personality and Social Psychology, 32,* 737–747. doi:10.1037/0022-3514.32.4.737

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–502. doi:10.1017/S0140525X00057046

Thagard, P. (1992). *Conceptual revolutions.* Princeton, NJ: Princeton University Press.

Thibaut, J. W., & Riecken, H. W. (1955). Some determinants and consequences of the perception of social causality. *Journal of Personality, 24,* 113–133. doi:10.1111/j.1467-6494.1955.tb01178.x

Thomas, R. P., Dougherty, M. R., Sprenger, A., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115,* 155–185. doi:10.1037/0033-295X.115.1.155

van Overwalle, F. (1998). Causal explanation as constraint satisfaction: A critique and a feedforward connectionist alternative. *Journal of Personality and Social Psychology, 74,* 312–328. doi:10.1037/0022-3514.74.2.312

van Overwalle, F., & van Rooy, D. (2001). How one cause discounts or augments another: A connectionist account of causal competition. *Personality and Social Psychology Bulletin, 27,* 1613–1626. doi:10.1177/01461672012712005

Wagenmakers, E. J. M., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology, 48,* 332–367. doi:10.1016/j.cogpsych.2003.08.001

Weber, E. U., Johnson, E., Milch, K., Chang, H., Brodscholl, J., & Goldstein, D. (2007). Asymmetric discounting in intertemporal choice: A query theory account. *Psychological Science, 18,* 516–523. doi:10.1111/j.1467-9280.2007.01932.x

Wegner, D. M. (2004). Precis of *The Illusion of Conscious Will. Behavioral and Brain Sciences, 27,* 649–659. doi:10.1017/S0140525X04000159

Weiner, B. (1985a). An attributional theory of achievement motivation and emotion. *Psychological Review, 92,* 548–573. doi:10.1037/0033-295X.92.4.548

Weiner, B. (1985b). Spontaneous causal thinking. *Psychological Bulletin, 97,* 74–84. doi:10.1037/0033-2909.97.1.74

Wellman, M. P., & Henrion, M. (1993). Explaining "explaining away." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 287–292. doi:10.1109/34.204911

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56,* 161–169. doi:10.1037/0022-3514.56.2.161

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy–attribution account of feelings of familiarity. *Acta Psychologica, 98,* 141–165. doi:10.1016/S0001-6918(97)00040-1

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy–attribution hypothesis I: The heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 3–13. doi:10.1037/0278-7393.27.1.3

Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy–attribution hypothesis II: Expectation, uncertainty, surprise and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 14–33. doi:10.1037/0278-7393.27.1.14

Wilson, D., & Sperber, D. (2004). Relevance theory. In G. Ward & L. Horn

(Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford, United Kingdom: Blackwell Science.

Wimer, S., & Kelley, H. H. (1982). An investigation of the dimensions of causal attribution. *Journal of Personality and Social Psychology, 43,* 1142–1162. doi:10.1037/0022-3514.43.6.1142

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136,* 82–111. doi:10.1037/0096-3445.136.1.82

Wolff, P., Barbey, A., & Hausknecht, M. (2010). For want of a nail: How absences cause events. *Journal of Experimental Psychology: General, 139,* 191–221. doi:10.1037/a0018129

Zuckerman, M., Eghrari, H., & Lambrecht, M. R. (1986). Attributions as inferences and explanations: Conjunction effects. *Journal of Personality and Social Psychology, 51,* 1144–1153. doi:10.1037/0022-3514.51.6.1144

Received October 1, 2009
Revision received June 15, 2010
Accepted September 16, 2010 ■