ELSEVIER

Cognitive Science 25 (2001) 565–610

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Naive causality: a mental model theory of causal meaning and reasoning

Eugenia Goldvarg[a], P.N. Johnson-Laird[b,*]

[a]*Infant Cognition Laboratory, Department of Brain and Cognitive Science, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

[b]*Department of Psychology, Princeton University, Green Hall, Princeton, NJ 08544, USA*

All reasonings concerning matter of fact seem to be founded on the relation of *Cause and Effect.*

David Hume, 1748/1988

Abstract

This paper outlines a theory and computer implementation of causal meanings and reasoning. The meanings depend on possibilities, and there are four weak causal relations: *A causes B, A prevents B, A allows B,* and *A allows not-B,* and two stronger relations of cause and prevention. Thus, *A causes B* corresponds to three possibilities: A and B, not-A and B, and not-A and not-B, with the temporal constraint that B does not precede A; and the stronger relation conveys only the first and last of these possibilities. Individuals represent these relations in mental models of what is true in the various possibilities. The theory predicts a number of phenomena, and, contrary to many accounts, it implies that the meaning of causation is not probabilistic, differs from the meaning of enabling conditions, and does not depend on causal powers or mechanisms. The theory also implies that causal deductions do not depend on schemas or rules.

1. Introduction

You think about causal relations, because they allow you to infer what will happen. Drinking too much wine will *cause* a hangover. Spraying yourself with insect repellent will *prevent* mosquitoes from biting you. Taking a short cut will *allow* you to avoid the traffic. Reasoning about matters of fact is, as Hume says, largely based on causal relations.

* Corresponding author. Te.: +1-609-258-4432; fax: +1-609-258-1113.
E-mail address: phil@princeton.edu (P.N. Johnson-Laird).

PII: S0364-0213(01)00046-5

Psychologists have studied thinking about causal relations, but they have yet to agree on a theory of the process (see e.g., White, 1995; Ahn & Bailenson, 1996; Hilton & Erb, 1996; Cheng, 1997). Like philosophers, they also disagree about the meaning of causal claims. Our concern is not with the philosophical problems of causation – with whether causes really exist, with whether, if they do exist, they are objective states in the world or subjective notions in the mind, or with whether they hold between facts, events, processes, or states of affairs (cf. Vendler, 1967). Our aim is to develop a psychological account of causal meaning and reasoning. We assume that causes can concern events, processes, and also states of affairs, for example, for want of a nail the shoe was lost. And so we will use the neutral expression, "states of affairs" or "states" for short, to embrace physical or psychological events, situations, facts, and other potential arguments of causal relations. We have no doubt that cause exists as an everyday concept, and we suspect that any sound philosophy of causation must reflect this fact. Otherwise, we pass over fine metaphysical distinctions, not because they are unimportant, but because they would take us too far afield from our main goals.

Our concern is with the everyday causal claims and, in particular, with what naive individuals, that is, those who have not studied philosophy or logic, take such claims to mean. And our goals are to formulate a psychological theory of the semantics of causation, to examine its consequences, and to support it with corroboratory evidence. The theory offers an account of how people think about causality, and it purports to solve three puzzles: first, what causal relations mean; second, how they are mentally represented; and, third, how people make inferences from them. These questions are logically prior to the problem of the induction of causal relations from observations. We need to know *what* is induced before we can properly understand *how* it is induced. Our theory also has implications for several common assumptions:

1. Causation is a probabilistic notion.
2. There is no semantic or logical distinction between causes and enabling conditions.
3. Inferences about causal relations depend on schemas or rules of inference.

If our theory is right, then each of these assumptions is wrong.

The theory is based on mental models. They were originally postulated in order to explain the comprehension of discourse (e.g., Stevenson, 1993; Garnham & Oakhill, 1996) and deductive reasoning (e.g., Johnson-Laird & Byrne, 1991). Mental models are the end result of perception, imagination, and the comprehension of discourse (Johnson-Laird, 1983). Each model corresponds to a possibility, and models are labeled to distinguish physical, deontic, and logical possibilities (Johnson-Laird & Byrne, 1991). Causation depends on physical possibilities. Its counterparts in the other two domains are obligation and logical necessity. How people determine the status of a possibility–whether it is physical, deontic, or logical–is an important epistemological problem. It need not detain us, however, since the crucial fact for us is merely that they do make these distinctions.

The structure and content of a model capture what is common to the different ways in which the possibility can occur. According to the theory, naive reasoners imagine the states of affairs described by premises: they construct mental models of them, and they establish the validity of an inference by checking whether its conclusion holds in these models

(Johnson-Laird & Byrne, 1991). The model theory offers a unified account of various sorts of reasoning. Thus, a conclusion is necessary–it *must* be the case–if it holds in all the models of the premises; it is possible–it *may* be the case–if it holds in at least one of the models; and its probability–assuming that each model is equiprobable–depends on the proportion of models in which it holds. However, in order to minimize the load on working memory, people represent as little information as possible. A fundamental twofold assumption is:

The principle of *truth*: i. The mental models of a set of assertions represent only those situations that are possible given the truth of the assertions. ii. Each such model represents what is described by clauses in the premises (affirmative or negative) only when they are true within the possibility.

The principle is subtle, and so we illustrate how it works with an example. An exclusive disjunction: There is a circle or else there is not a cross, calls for two mental models, one for each possibility, which we represent on separate lines:

o

¬ +

In accordance with the principle of truth, the first of these models represents explicitly that there is a circle, but it does not represent that in this possibility it is false that there is not a cross. Likewise, the second model represents explicitly that there is not a cross ("¬" denotes negation), but it does not represent that in this possibility it is false that there is a circle. The theory assumes that reasoners try to remember what is false, but that these "mental footnotes" soon tend to be forgotten. If individuals can keep track of the mental footnotes, they can envisage *fully explicit* models. Thus, the mental models of the preceding disjunction can be fleshed out to yield the fully explicit models:

o +

¬o ¬ +

Do models never represent what is false? That cannot be the case, because an important developmental milestone is the ability to pass the "false beliefs" test. Thus, four-year olds are usually able to distinguish their own beliefs from the false beliefs of another individual (see e.g., Leslie, 1994). Children are likewise able to engage in pretense, to detect lies and deception, and to reason on the basis of counterfactual suppositions. They may even be able to carry out Wason's (1966) selection task. In all these cases, as Denise Cummins (personal communication) pointed out, reasoners need to represent what is false. A crucial feature of false beliefs in studies of the "theory of mind" is that they do not depend on sentential connectives, that is, they are simple "atomic" beliefs (Alan Leslie, personal communication). It is straightforward to represent the falsity of such propositions, but it is much harder to envisage what corresponds to the falsity of an assertion containing several sentential connectives. Even adults do not have direct access to these possibilities but must infer them (see Johnson-Laird & Barres, 1994).

Experimental evidence has corroborated the model theory (see e.g., Johnson-Laird & Byrne, 1991). It accounts for reasoning about possibilities (Bell & Johnson-Laird, 1998), spatial and temporal relations (Byrne & Johnson-Laird, 1989; Schaeken et al., 1996);

counterfactual states (Byrne, 1997; McEleney & Byrne, 2001), probabilities (Johnson-Laird et al., 1999), and the consistency of assertions (Johnson-Laird et al., 2000). Perhaps the best evidence for the theory derives from its prediction of egregious, but systematic, errors (see e.g., Johnson-Laird & Savary, 1999; Yang & Johnson-Laird, 2000). We return to these "illusory" inferences in Section 4.

To explain tasks going beyond deduction, we must make additional assumptions. In the present paper, we frame a new theory based on assumptions about causation. The paper begins with this theory and its computer implementation (Section 2). It reports studies of causal meanings that corroborate the theory (Section 3); and it reports studies of causal deduction that also support it (Section 4). Finally, it considers some implications of the theory and shows how they dispel several common misconceptions about causation (Section 5).

2. The model theory of causal meanings and reasoning

Our aim is to outline a theory of the meaning of causal relations in everyday life and of how people reason with them. It is important, however, to distinguish three sorts of assertion: *General* causal assertions such as:

Heating pieces of metal causes them to expand;

Singular causal assertions where the outcome is known, such as:

Heating this piece of metal caused it to expand;

and *Singular* causal assertions where the outcome is not known, such as:

Heating this piece of metal will cause it to expand.

Hume (1748/1988, p. 115) bequeathed us two definitions of the meaning of causal assertions: "We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second.* Or in other words, *where, if the first object had not been, the second never had existed.*" According to his first definition, the general causal assertion above has the following construal: For any x, if x is a piece of metal and x is heated then x expands. Similarly, a singular cause where the outcome is not known has the construal: If this piece of metal is heated then it will expand. According to Hume's second definition, a singular cause where the outcome is known has a counterfactual construal: If this piece of metal had not been heated then it would not have expanded.

This section begins with the model theory of causal assertions, and contrasts it with other current theories, including probabilistic accounts and theories that do not distinguish the meanings of causes and enabling conditions. It outlines the difference between mental models and fully explicit models of causal relations. It describes a computer program implementing the theory. It shows how causal reasoning can be based on models. Finally, it reviews the principal predictions of the theory.

2.1. The theory of meaning

There are at least five putative components of causation: temporal order, spatial contiguity, necessary connection, probabilistic connection, and causal powers or mechanisms.

Philosophers have long discussed these components (e.g., Hume, 1739/1978; Carnap, 1966; Lewis, 1973; and Mackie, 1980), and we consider each of them in order to develop the model theory.

2.1.1. Temporal order

In daily life, the normal constraint on a causal relation between A and B is that B does not precede A in time (see e.g., Tversky & Kahneman, 1980; Bullock et al., 1982). Hence, the model theory postulates:

The principle of *temporal constraint:* given two states of affairs, A and B, if A has a causal influence on B, then B does not precede A in time.

This principle allows that a cause can be contemporaneous with its effect–squeezing the toothpaste tube, as Paolo Legrenzi noted (personal communication), can be contemporaneous with the toothpaste coming out of the tube (see Kant, 1781/1934; Taylor, 1966). Teleological claims–the need to get toothpaste caused the squeezing of the tube–do not violate the principle granted that a representation of the goal caused the behavior (Daniel Schwartz, personal communication). Hume (1739/1978) argued that the cause precedes the effect, and rejected contemporaneity. He finally declared that the matter was of no great importance. A causal interpretation of Newton's theory of gravitation, however, calls for instantaneous effects at a distance. And individuals violate even our weaker constraint when in discussing time travel, they assert that an event in the present caused an event in the past. Physicists and philosophers have likewise speculated about violations of the temporal constraint. For this reason, we treat it as true for everyday causal assertions, but as deliberately violated in certain sorts of discourse.

2.1.2. Spatial contiguity

There may be innate constraints on the perception of causal relations as when you see one object bump into another and cause it to move (Michotte, 1946/1963). This claim is corroborated by studies of infants' perception (e.g., Leslie, 1984). It has led Geminiani et al. (1996), echoing Hume (1739/1978), to argue that adult conceptions of physical causation depend ultimately on physical contact. Likewise, Lewis (1986) showed that three principles suffice for a computer program that constructs models of a physical system: every event has a cause; causes precede their effects; and an action on an object is likely to be the cause of any change in it. Yet, the concept of causation in daily life is much broader and more abstract than the perception of cause in simple Michotte-like mechanical events (Miller & Johnson-Laird, 1976). People often make and understand causal assertions that violate the constraint of spatial contiguity. They may refer to "action at a distance" in remarks about physical states, such as: The moon causes the tides, and in statements about psychological states: Pat's rudeness caused Viv to get annoyed. We accordingly assume that spatial contiguity–physical contact–is not a necessary component of the everyday *meaning* of causal relations. But, it may underlie many of the inferences that individuals make to explain such claims.

2.1.3. Necessary connection

According to the model theory, the central components of the meanings of causal relations are the possibilities to which the relations refer:

The *modal* principle of causation: given two states of affairs, A and B, the meaning of a causal relation between them concerns what is possible and what is impossible in their co-occurrences (i.e., "modalities").

This principle implies a necessary connection between cause and effect. Hume rejected such a connection in favor of constant conjunction; Kant (1781/1934) restored it, arguing in effect that it was a component of an innate conception. In everyday life, when we assert *A will cause B,* we mean that if A occurs then B must occur. Hence, necessity is part of the meaning of such claims (see Harré & Madden, 1975).

Following the principle of causal modalities, a singular causal claim of the form *A will cause B* means that there are the following possibilities satisfying the temporal constraint:

1. a b
   ¬a b
   ¬ ¬b

Each row denotes a model of an alternative possibility. Thus, the models represent the possibility of A with B, and possibilities in which A does not occur, either with or without B. The assertion is false if A occurs without B. A general causal claim, *A causes B,* calls for just a single model of various possible events within the same situation or "universe of discourse":

a b
¬a b
¬a ¬b
¬a ¬b

Each row represents a possibility, and the various possibilities may co-occur in the situation. The assertion is false if there is a possibility in which A occurs without B. A singular causal assertion where the outcome is known, *A caused B,* has a model of the factual situation, and models representing alternative, but counterfactual, possibilities:

a b (the factual case)
¬a b (a counterfactual possibility)
¬a ¬b (a counterfactual possibility)

A counterfactual possibility is a state that was once was a real possibility but that did not, in fact, occur (Johnson-Laird & Byrne, 1991). The falsity of *A caused B* is slightly more complicated than the previous cases. It is false, of course, if either A or B did not occur. But, it can be false even if both A and B occurred, because the relation between them was not causal. Here the counterfactual possibilities are critical. If they include a case in which A occurred without B, or in which A occurred after B, then the assertion is false. These factual and counterfactual representations are consistent with Cheng's (1997) concept of a "focal" set.

There are a variety of other causal relations, and each of them can occur in general and singular assertions. For simplicity, we focus on singular causal relations with unknown outcomes, and we take for granted the temporal constraint. An assertion of the form *A will prevent B* means that there are the following possibilities:

2. a ¬b

¬a b

¬a ¬b

If the assertion is false, then there is a possibility in which A occurs with B.

An assertion of the form *A will allow B* such as: Taking the short cut will allow you to avoid the traffic, has a strong *implicature* that not taking the short cut will not allow you to avoid the traffic. An implicature is an inference that is warranted by pragmatic considerations (see Grice, 1975; and for an account of how pragmatics is implemented within the model theory, see Johnson-Laird & Byrne, 2001). In this case, individuals who speak in an informative way would not make the claim if they knew that one could just as well avoid the traffic by not taking the short cut. Thus, an assertion of the form *A will allow B* means either that all four contingencies are possible or that there are the following three possibilities satisfying the implicature:

3. a b

a ¬b

¬a ¬b

The assertion is false if there is a possibility in which B occurs in the absence of A. The verb *allow* is ambiguous in that it also has a sense of giving permission. Some authors accordingly prefer to use the verb *enable* and to talk of "enabling conditions." Unfortunately, this verb too can have the same ambiguity (cf. "The Royal toast enabled us to smoke"). In fact, the ambiguity appears to arise from the fact that both senses can be construed as *to make possible*, where in one case the possibility concerns physical matters and in the other case it concerns deontic matters, that is, what is permissible. In general, we will use "allow" because "enable" has a connotation that the result was intended. It is odd to assert, for example: Shoddy work enabled the house to collapse.

An assertion of the form *A will allow not B* means either that all four contingencies are possible or that there are the following three possibilities satisfying the implicature:

4. a ¬b

a b

¬a b

It is sometimes important to distinguish between meaning and implicature, but in what follows we will not attempt to keep them apart.

The preceding relations are weak, for example, given *A will cause B*, A is sufficient for B, but not necessary, because B may have other causes. Indeed, if one ignores the temporal constraint, the models of causal possibilities capture the notion of *sufficiency* and the models of enabling possibilities capture the notion of *necessity* as invoked by various theorists (e.g.,

Thompson, 1995). An example of weak causation is: Eating too much will cause you to gain weight. You may also gain weight as a result of failing to exercise. In addition to the four weak relations, there are two strong relations. *A, and only A, will cause B* means that there are the following possibilities satisfying the temporal constraint:

5. a b
   ¬a ¬b

In other words, A is the unique cause and enabler of B. And *A, and only A, will prevent B* means that there are the following possibilities satisfying the temporal constraint:

6. a ¬b
   ¬a b

A case of strong causation is: Drinking too much alcohol will cause you to get drunk. Drunkenness has no other cause. The models for the six causal relations are summarized in Table 2 below.

There is no causal relation if A or B is bound to occur or bound not to occur; and Cheng & Nisbett (1993) have shown that naive individuals concur with this claim. An assertion such as: "The rotation of the earth causes the prevailing wind patterns," might seem like a counterexample on the grounds that the earth is bound to move. In fact, naïve individuals readily envisage the counterfactual possibility that the earth does not rotate, because they assent to the claim: If the earth were not to rotate then there might be no prevailing wind patterns. If all you know is that B is bound to occur if A occurs, then you are not entitled to make any causal claim about the relation between them. On the one hand, B may also be bound to occur even if A does not occur, that is, there are only the following possibilities:

a b
¬a b

and so it would be wrong to invoke a causal relation between A and B. On the other hand, even granted a causal relation between the two, A could be a strong or a weak cause of B. The existence of several sorts of causal relation may come as a surprise to the reader. Philosophers have often seemed to assume that there is only a single relation of cause and effect, and, according to Lombard (1990), their neglect of enabling conditions has led them into error.

The general weak relations depend on the temporal constraint and quantification over possible states, that is, the relations are not merely truth-functional:

1. *A causes B*: for any possibility in which A occurs B occurs.
2. *A prevents B*: for any possibility in which A occurs B does not occur.
3. *A allows B*: there is at least one possibility in which A occurs and B occurs.
4. *A allows not B*: there is at least one possibility in which A occurs and B does not occur.

The assertion: For any possibility in which A occurs B occurs, is logically equivalent to: There is no possibility in which A occurs and B does not occur. Of course, individuals often assert general causal claims, such as: Smoking causes cancer, when they know that coun-

terexamples occur. This phenomenon has led some theorists to invoke a probabilistic analysis of causation–an analysis that we consider below.

Our account implies that there should be many ways to describe causal relations. This variety has advantages. Three possibilities are hard to hold in mind at the same time, but one description can focus on one possibility and another description focus on another possibility:

1.	a	b	A causes B
	¬a	b	
	¬a	¬b	Not-A allows not-B
2.	a	¬b	A prevents B
	¬a	b	Not-A allows B
	¬a	¬b	
3.	a	b	A allows B
	a	¬b	
	¬a	¬b	Not-A prevents B
4.	a	b	
	a	¬b	A allows not-B
	¬a	b	Not-A causes B

The same analysis applies to the large set of verbs that express specific causal relations, such as *annoy*, *kill*, and *break* (see Miller & Johnson-Laird, 1976, Sec. 6.3).

2.1.4. Probabilistic connection

In contrast to the model theory's modal account, a twentieth century view in philosophy and psychology is that causation is a probabilistic notion. Reichenbach (1956) proposed such an analysis, arguing that *C causes E* if:

$$p(E|C) > p(E|\neg C)$$

where $p(E|C)$ denotes the conditional probability of the effect E given that the putative cause C occurs, and $p(E|\neg C)$ denotes the probability of the effect given that the cause does *not* occur. He allowed, however, that a putative cause could be "screened off" if there was an earlier probabilistic cause, D, yielding both C and E, that is, $p(E|C) = p(E|D)$, and $p(E|\neg C) = p(E|\neg D)$. Suppes (1970) defended a similar analysis, even for causation in ordinary experience (see also Suppes, 1984, p. 54; and Salmon, e.g., 1980, for a defense of a more sophisticated theory).

A psychological theory of probabilistic causes has been defended by Cheng and her colleagues (e.g., Cheng & Novick, 1990; see also Schustack, 1988). There are also parallel probabilistic accounts of the meaning of conditionals (see e.g., Newstead et al., 1997). "Because causal relations are neither observable nor deducible," Cheng (1997, p. 367) writes, "they must be induced from observable events." She assumes that one observes the differ-

ence in the respective conditional probabilities shown in the inequality above. The probabilities can be computed from a partition of the events, such as the following one stated with frequencies of occurrence:

C	E	28
C	¬E	3
¬C	E	10
¬C	¬E	59

The difference between p(E | C) and p (E | ¬ C) is known as the probabilistic *contrast.* When it is noticeably positive (as above: 0.76), *C causes E;* when it is noticeably negative, *C prevents E.* The contrast model fails to make the correct predictions for certain causal inductions, and so Cheng (1997) has proposed a "power probabilistic contrast" model (the Power PC model) in which the contrast is normalized by dividing it by the base rate for the effect, that is, 1 – P(E | ¬ C).

The main evidence for a probabilistic semantics is that people judge that a causal relation holds in cases, such as the partition above, in which the antecedent is neither necessary nor sufficient to bring about the effect (e.g., McArthur, 1972; Cheng & Novick, 1990; Cummins et al., 1991). One might therefore suppose that causal relations are intrinsically probabilistic. Certainly, people often *induce* causal relations from probabilistic data. Yet, our hypothesis is that the meaning of a causal relation is not probabilistic, though the evidence supporting it may be probabilistic. We return to this hypothesis later in the light of our experimental results.

2.1.5. Causal powers or mechanisms

Theorists often invoke some causal power, mechanism, or means of production, over and above the relevant possibilities (e.g., Harré & Madden, 1975; Ahn et al., 1995; White, 1995; and Cheng, 1997). One candidate might be a framework of scientific laws or explanatory principles (Carnap, 1966). These principles, Hart & Honoré (1985) argue, lie behind every causal assertion about singular events. Another candidate is that "a causal mechanism is the process by which a cause brings about an effect" (Koslowski, 1996, p. 6). Still another candidate is the notion that causes produce effects through a transmission of energy or some other property from cause to effect. "In this sense," Harré & Madden (1975, p. 5) write, "causation always involves a material particular which produces or generates something." When a hammer is used to smash a plate, for example, a specific causal power of the hammer blow produces the effect: the energy in the blow is transferred to the plate, causing it to break. Theories of causal power accordingly define causation in terms of the intrinsic properties of objects. Hume (1739/1978), however, rejected intrinsic properties that produce effects, arguing that the definition of *production* is indistinguishable from that of causation itself.

The principal evidence for causal power comes from studies showing that beliefs can override information about the covariation of cause and effect (e.g., White, 1995; Koslowski, 1996). The phenomena, however, do not call for the introduction of causal properties,

powers, or mechanisms, into the *meaning* of causal relations. People make causal assertions when they have no idea of the underlying causal mechanism, for example: Smoking opium causes people to fall asleep. Hence, the meaning of many assertions would imply at most only that some unknown power exists: *A causes B* means that there exists a mechanism or power by which A causes B. But, it would then be necessary for theorists to specify what counts as a causal mechanism or power, or else the existential clause in this analysis is vacuous. To the best of our knowledge, as Hume anticipated, no-one has succeeded in formulating a satisfactory definition of a causal power or mechanism, which itself makes no reference to causality. Our skepticism, however, in no way impugns the role of known causal mechanisms in the induction of causal relations–a topic to which we return in the General Discussion.

2.2. Circumstances and how causes differ in meaning from enabling conditions

The model theory distinguishes between causing an effect and allowing it to occur, that is, between causes and enabling conditions. Yet, current psychological theories deny that there is any logical or semantic distinction between the two. The argument, which is originally due to Mill (1874), influenced philosophers and jurists first, and then psychologists. It can be illustrated by an example of a singular cause. Consider an explosion that is caused by the occurrence of a spark in a container of combustible vapor. The explosion would not have occurred in the absence of the spark, and it would not have occurred in the absence of the vapor. Hence, both the spark and the vapor are individually necessary and jointly sufficient to cause the explosion. Yet, people often speak of the spark as the *cause* of the explosion, and the presence of the vapor as the *enabling* condition that allows it to occur. Likewise, the absence of the spark or the vapor would be a *disabling* condition that would not allow the explosion to occur (see Cummins, 1995, 1998). If the difference between causes and enabling conditions cannot be accounted for by logic or meaning, then what distinguishes the two? Mill himself thought that the choice of a cause was often capricious, but he did offer some more systematic answers, as have many other authors.

According to one school of thought, causes are abnormal whereas enabling conditions are normal. Thus, Hart & Honoré (1959/1985) argued that when individuals identify single causes, they choose the unusual factor as the cause–the spark rather than the vapor in the example above, or else they choose a voluntary human action. Girotto et al., (1991) have independently discovered that voluntary human actions are the main events that reasoners seek to undo in thinking counterfactually about the causes of unfortunate events. Various ways exist to try to determine what is unusual. One can distinguish what is rare and what is common (Hart & Honoré, 1959/1985), what is inconstant and what is constant (Cheng & Novick, 1991), or what does and does not violate a norm that can be assumed by default (see e.g., Kahneman & Tversky, 1982; Kahneman & Miller, 1986; Einhorn & Hogarth, 1986).

According to another school of thought, the cause is the factor that is conversationally relevant in explanations (Mackie, 1980; Turnbull & Slugoski, 1988). Thus, Hilton & Erb (1996) argue for a two stage process: “explanations are first cognitively generated by building mental models of the causal structure of events, from which particular factors are identified in conversationally given explanations” (p. 275). These authors adduce Grice’s (1975) maxims of conversation as determining the cause, but they allow that other accounts

Table 1
A summary of hypotheses about the difference between causes and enabling conditions. All these accounts presuppose that there is no difference in meaning or logic between the two

Distinguishing characteristics		Examples of proponents
Causes	Enabling conditions	
Most recent event	Earlier event	Mill (1874)
Rare event	Common event	Hart and Honoré (1985)
Inconstant state	Constant state	Cheng and Novick (1991)
Violates norm	Conforms to norm	Einhorn and Hogarth (1978)
Relevant to conversation	Not relevant	Mackie (1980)
		Turnbull and Slugoski (1988)
		Hilton and Erb (1996)

of pragmatics would serve as well. Yet another school of thought supposes that it is the most recent or precipitating event that is the cause (Mill, 1874), and there are still other views (see Hesslow, 1988, for a review). Table 1 summarizes the main hypotheses about the distinction between causes and enabling conditions. Of course, all these components might be true, and yet we will show that there *is* a difference in meaning too.

To recapitulate the argument: the model theory entails that a distinction in meaning and logic exists between causing and allowing. Other psychological theories presuppose that the two are not logically distinct. Hence, an apparent paradox needs to be resolved. We need to explain why Mill's argument seems so compelling. In fact, the answer comes from the subtle effects of knowledge in determining the *circumstances* of a cause and effect.

Suppose you observe the following sequence of events: A doctor injects a patient with a drug, and then the patient loses consciousness. What is the causal relation, if any, between the two events? The observation is inconsistent with two causal relations: the injection did not prevent loss of consciousness in either its strong or weak sense. But, it is compatible with any of the four remaining relations and, of course, with the lack of any causal relation at all. A corollary of this uncertainty is that the mere observation of a particular sequence of states–in the absence of knowledge about causal relations–never suffices to establish a unique causal relation. Even the simple perception of physical causes, as in Michotte's (1963) studies, is not a counterexample. The participants saw a causal relation when one object collided with another and launched it into motion. They were mistaken of course, because Michotte's apparatus did not use real physical objects colliding with one another. Causal relations are modal. They are not merely about what occurred but also about what might have occurred. What might have occurred, however, cannot be determined from observation unsupported by knowledge of the circumstances. The model theory accordingly postulates:

> The principle of *circumstantial interpretation*: Causal interpretation depends on how people conceive the circumstances of states, that is, on the particular states that they consider to be possible, whether real, hypothetical, or counterfactual.

Precursors to this idea include Hart & Honoré's (1959/1985) and McGill's (1989) "context" of a cause, Mackie's (1980) "causal field," and Cheng & Novick's (1991) "focal set" of events. The circumstantial principle, however, implies that you use your general knowledge

and your knowledge of the state at issue to generate models. In the case of a singular causal claim, one model represents the actual state, and the other models represent the relevant possibilities in the circumstances. The models fix the appropriate causal relation.

To return to the preceding example, if the circumstances are as follows:

injection loss-of-consciousness

¬injection loss-of-consciousness

¬injection ¬loss-of-consciousness

an appropriate description is: The injection caused the patient to lose consciousness. If the circumstances are as follows:

injection loss-of-consciousness

injection ¬loss-of-consciousness

¬injection ¬loss-of-consciousness

an appropriate description is: The injection allowed the patient to lose consciousness. And if the circumstances are as follows:

injection loss-of-consciousness

injection ¬loss-of-consciousness

¬injection loss-of-consciousness

an appropriate description is: The injection did not prevent the patient from losing consciousness.

The truth of an assertion about past states of affairs, such as: The injection caused the patient to lose consciousness, presupposes that the patient had an injection and that the patient lost consciousness. But, a causal relation is not true merely because both propositions that it interrelates are true (Burks, 1951; Mackie, 1980). Even granted the temporal constraint on the order of the two events, the modal nature of causation is borne out by its support for counterfactual claims (see Hume, 1748/1988; Mill, 1874; and many recent accounts). A corollary is that the appropriate description of an observation depends on circumstantial interpretation. Consider the circumstances of a strong causal relation, such as:

injection loss-of-consciousness

¬injection ¬loss-of-consciousness

With an observation corresponding to the first of these models, the appropriate counterfactual assertion is: If the patient hadn't had the injection, then he wouldn't have lost consciousness. With an observation corresponding to the second of the models, the appropriate counterfactual assertion is: If the patient had had the injection, then he would have lost consciousness. Counterfactual descriptions are not unique to singular causal claims, as an example from Lewis (1973) illustrates: If kangaroos had no tails then they would topple over, is just another way of expressing the general claim: Kangaroos' tails prevent them from toppling over.

What are the correct circumstances of a state? Well, it all depends. Often, there are no decisive criteria. That is why the circumstances–the *ceteris paribus* clause of counterfactual conditionals–have bedeviled philosophical analyses (e.g., Stalnaker, 1968). They play an important role in blocking otherwise valid inferences of the following sort (see e.g., Lewis, 1973):

If the match had been struck, then it would have lighted.

∴ If the match had been soaked in water and struck, then it would have lighted.

The circumstances of the conclusion are no longer those of the premises. The case is altered; the conclusion is false even if the premise is true (see Johnson-Laird & Byrne, 2001, for an account of how general knowledge modulates the interpretation of such conditionals).

We can now consider again the example illustrating Mill's argument. A spark in a combustible vapor causes an explosion. You know that in the presence of the vapor, the spark causes the explosion, and that in the absence of either the vapor or the spark, there is no explosion. Your knowledge accordingly yields the circumstances shown in the following models:

vapor	spark	explosion
vapor	¬spark	¬explosion
¬vapor	spark	¬explosion
¬vapor	¬spark	¬explosion

The roles of the spark and vapor *are* equivalent. Jointly, they are the strong cause of the explosion.

You can envisage other circumstances. Suppose, for example, that a tank is used to store gasoline, and at present it may or may not contain a combustible vapor. The presence of the vapor will allow an explosion to occur, and the occurrence of a spark, or, say, a naked flame will cause an explosion. The circumstances are shown in the following models:

vapor	spark	explosion
vapor	¬spark	explosion
vapor	¬spark	¬explosion
¬vapor	spark	¬explosion
¬vapor	¬spark	¬explosion

The respective roles of the two antecedents are logically distinct. The possible combinations of vapor and explosion in the set of models are as follows:

vapor	explosion
vapor	¬explosion
¬vapor	¬explosion

They show that the vapor allows the explosion to occur. The possible combinations of spark and explosion in the set of models are as follows:

spark	explosion
spark	¬explosion
¬spark	explosion
¬spark	¬explosion

The full set of circumstances, however, shows that given the presence of the vapor, the spark causes the explosion to occur.

You can envisage circumstances in which the causal roles of the vapor and spark are interchanged. Suppose, for example, that an induction coil delivers a spark from time to time in an enclosed canister. You know that the introduction of a combustible vapor will cause an explosion. You also know that the occurrence of the spark allows such an explosion to occur. It may even occur without the vapor if, say, an explosive substance such as gunpowder is put into the canister. The circumstances are as follows:

spark	vapor	explosion
spark	¬vapor	explosion
spark	¬vapor	¬explosion
¬spark	vapor	¬explosion
¬spark	¬vapor	¬explosion

The spark allows the explosion to occur, and given its occurrence, the vapor causes the explosion.

This analysis shows that causes and enabling conditions are distinct, and that they reflect the modal and circumstantial principles. The respective probabilities of each of the possibilities can make one antecedent common and the other antecedent rare, but the probabilities have no bearing on their causal roles (*pace* Hart & Honoré, 1959/1985). Indeed, the switch in causal roles from one set of circumstances to the other shows that none of the factors in Table 1 is essential to distinguishing between causes and enabling conditions. Causes need not be unusual or abnormal, and they need not be pragmatically relevant to explanations. Cheng & Novick's (1991) experiments corroborated both these claims, and their theory of causal induction also relies on sets of possibilities. But, they argued that enabling conditions are *constant* in a set of possibilities whereas causes are not constant (see also Cheng, 1997). It is true that enabling conditions are constant in some circumstances. But, our preceding examples show that constancy in the circumstances is not necessary for an enabling condition. Neither the spark nor the vapor is constant in the circumstances above, yet their logical roles are distinct, and one is an enabling condition for the other to function as a cause.

2.3. Mental models of causal relations

Our preceding analysis of the meanings of causal relations is in terms of fully explicit models. The principle of truth, however, predicts that naive individuals will tend to rely on the corresponding *mental* models for each of the causal relations. An assertion of the form: *A will cause B,* calls for the mental models:

a b

. . .

in which the ellipsis represents implicitly possibilities in which the antecedent, A, does not hold. There is a mental footnote to capture this fact–in effect, A cannot occur in the possibilities represented by the implicit model. Given the mental footnote, it is possible to flesh out the models fully explicitly as:

a b

¬a b

¬a ¬b

The theory postulates that individuals normally reason on the basis of mental models, but with simple assertions of the present sort, they can appreciate that *A will cause B* is compatible with B having other causes. Given time, they may even enumerate all three explicit possibilities. Strong causation as expressed by *A, and only A, will cause B* has the same mental models as weak causation, but the mental footnote indicates that neither A nor B can occur in the possibilities represented by the implicit model, and so the only way to flesh out the model fully explicitly is as shown below:

a b

¬a ¬b

An assertion of the form: *A will prevent B,* relies on a negative verb (Clark & Clark, 1977), and so it means that if A occurs then B does not occur in the circumstances:

a ¬b

. . .

where there is a mental footnote indicating that A cannot occur in the possibilities represented by the implicit model. Again, individuals may be able to enumerate the explicit possibilities, but normally they should use mental models to think about the relation. Analogous considerations apply to assertions of the form: *A allows B,* and *A allows not B.*

2.4. The computational implementation of the theory

We have developed a computer program of the model theory. Its input is a set of causal assertions, and it constructs the corresponding set of mental models and, for purposes of comparison, the set of fully explicit models. It also draws the causal conclusion, if any, that

Table 2
The models for the six singular causal relations with unknown outcomes. The central column shows the mental models normally used by human reasoners, and the right-hand column shows the fully explicit models, which represent the false components of the true cases using negations that are true: "¬" denotes negation and ". . ." denotes a wholly implicit possibility. The mental models for the strong and weak relations of cause and prevention differ only in their mental footnotes (see text).

Connective	Mental models	Fully Explicit models
1. A will cause B:	A B . . .	A B ¬A B ¬A ¬B
2. A will prevent B:	A ¬B . . .	A ¬B ¬A B ¬A ¬B
3. A will allow B:	A B . . .	A B A ¬B ¬A ¬B
4. A will allow not-B:	A ¬B . . .	A ¬B A B ¬A B
5. A and only A will cause B:	A B . . .	A B ¬A ¬B
6. A and only A will prevent B:	A ¬B . . .	A ¬B ¬A B

the resulting models support. The program uses a grammar to parse each input sentence, and it relies on a lexical semantics for causal verbs, and a compositional semantics with semantic rules for each rule in the grammar to assemble the models. The models for each of the six singular causal relations with unknown outcomes are summarized in Table 2. When the outcome is known, the remaining models represent counterfactual states. The models in the table also stand for general causal assertions, but in this case each row represents an alternative possibility within the same situation.

The program conjoins sets of models according to the procedures summarized in Table 3. These procedures combine separate possibilities, either those corresponding to models of singular causal assertions or those in models of general causal assertions. Readers will recall that for general assertions, the theory postulates that reasoners construct a single model representing the different sorts of possibilities. We make no strong claims that reasoners form separate models and then combine them; they might instead add information from a premise to an existing model or set of models (see e.g., Bucciarelli & Johnson-Laird, 1999). In either case, however, the process yields the same end results.

Different sorts of model support different sorts of conclusions. In particular, the program uses a single model of the possibilities:

a c

. . .

to draw the general conclusion, *A causes C*. From a single model of the form:

Table 3
The procedures for forming a conjunction of two sets of possibilities. The procedures apply either to individual models (based on singular causal relations) or to individual possibilities (based on general causal relations). Each procedure is presented with an accompanying example. In principle, each procedure should take into account mental footnotes, but reasoners soon forget them. The program implementing the theory also reasons at more advanced levels, first taking footnotes into account, and then constructing fully explicit models.

1. For a pair of explicit items, the result conjoins their elements, and drops any duplicates:
a b and b c yield a b c
2. For a pair of items that contain an element and its contradiction, the result is null (akin to the empty set):
a b and ¬b c yield null
3. For null combined with any item, the result is null:
null and a b yield null
4. For a pair of implicit items, the result is implicit:
. . . and . . . yield . . .
5. For an implicit item combined with an explicit one, the result by default is null:
. . . and b c yield null
But, if the explicit item contains no element in common with anything in the same set from which the implicit item is drawn, then the result is the explicit item:
. . . and b c yield b c

a c

a

. . .

it draws the conclusion, *A allows C.* The program draws negative conclusions from two sorts of models, which both must contain negative tokens, either of an end term or a middle term. The model:

a ¬c

. . .

and the model:

a

c

. . .

yield the conclusion, *A prevents C.* Likewise, the model:

a ¬c

a

. . .

and the model:

a c

a

. . .

yield the conclusion, *A allows not-C.*

We illustrate how the program works by elucidating the contrast between causal forks and causal chains. Fisher (1959) argued that smoking might not be a cause of lung cancer, and that instead there could be an unknown gene, X, that both causes people to smoke and independently causes them to get lung cancer. If they have the deadly gene then they are likely to develop cancer whether or not they smoke. If they do not have the deadly gene then they are unlikely to develop cancer whether or not they smoke. Ergo, they might as well smoke. A contrasting view is that a causal chain occurs: the gene causes smoking, and smoking causes cancer. Given the assertions for a causal fork:

Gene causes smoking.

Gene causes cancer.

the program constructs the following fully explicit model:

Gene	Smoking	Cancer
¬Gene	Smoking	Cancer
¬Gene	Smoking	¬Cancer
¬Gene	¬Smoking	Cancer
¬Gene	¬Smoking	¬Cancer

The causal fork allows a possibility in which smoking occurs without cancer. In contrast, given the assertions for a chain of weak causes:

Gene causes smoking.

Smoking causes cancer.

the program constructs the following set of fully explicit models:

Gene	Smoking	Cancer
¬Gene	Smoking	Cancer
¬Gene	¬Smoking	Cancer
¬Gene	¬Smoking	¬Cancer

Both the gene and smoking cause cancer.

2.5. Deductive inferences from causal relations

Some theorists suppose that causes cannot be deduced, but only induced. But, the causal status of an observation *can* be deduced from knowledge of its circumstances. For example, suppose you know that a certain anesthetic causes a loss of consciousness, and you observe that a patient who is given the anesthetic loses consciousness. You can deduce that the

anesthetic *caused* the patient to lose consciousness. How do you make such deductions? There are three potential answers.

The first answer is that you rely on formal rules of inference (see e.g., Braine & O'Brien, 1998). One such account (Osherson, 1974-6) includes a fragment of modal logic, and Rips (1994, p. 336) has suggested that formal rules might be extended to deal with causal reasoning. Likewise, philosophers have proposed axiomatic systems for causal deductions (e.g., von Wright, 1973). Hence, causal deductions could depend on axioms (or "meaning postulates") of the following sort: *If X causes Y, and Y prevents Z, then X prevents Z*, where X, Y, and Z, are variables that take states as their values.

The second answer is that you make causal inferences relying on *pragmatic reasoning schemas* (e.g., Cheng et al., 1986). For instance, the preceding axiom is framed instead as a rule of inference:

X causes Y.

Y prevents Z.

∴ X prevents Z.

This rule and others make up a pragmatic schema, and causal reasoning as a whole could depend on several such schemas. The idea goes back to Kelley's (1973) theory of causal attribution, which postulated such schemas for checking causal relations. Morris & Nisbett (1993), for example, formulated a schema that includes the following two rules:

If cause A is present then effect B occurs.

Cause A is present.

∴ Effect B occurs.

and:

If cause A is present then effect B occurs.

Effect B does not occur.

∴ Cause A is not present.

Inferences about causation tend to be easier than equivalent inferences based on conditionals, at least in Wason's selection task (e.g., Markovits & Savary, 1992). Hence, Morris & Nisbett (1993) argue that schemas guide causal deduction, and that this claim is supported by the fact that graduate students in psychology improve in causal reasoning whereas those in philosophy do not. These phenomena, however, are open to alternative explanations–content and experience may, for example, enable individuals to flesh out their models of possibilities more explicitly (cf. Green & Over, 2000).

The third answer is that naive reasoners make causal deductions, not by using rules or schemas, but by constructing mental models. Given the appropriate temporal constraints and, say, the following mental model of possibilities:

a	b	c
¬a	¬b	c
¬a	b	¬c

they can validly infer: *A causes C*, because the model contains each of the possibilities required for this relation. In practice, naive individuals may not construct fully explicit models. The theory accordingly postulates:

The principle of *causal deduction:*

Causal deductions are based on models of the premises, and with complex premises, these models will tend to be mental models rather than fully explicit models. Granted the appropriate temporal constraints, a conclusion is valid if the set of premise possibilities corresponds to the set for the conclusion.

2.6. The predictions of the theory

The model theory makes three main empirical predictions. First, it predicts that the meanings of causal relations are modal, and so naïve individuals should consider that each sort of causal relation rules out certain states of affairs as impossible. In contrast, probabilistic accounts of the meaning of, say, *A will cause B,* are compatible with any possibility, because only the relative frequencies of possibilities matter. Second, the model theory predicts that causes and enabling conditions differ in meaning. Hence, naïve individuals should judge that different possibilities are compatible with assertions of the form, *A will cause B* and *A will allow B* (see Table 2). Likewise, they should draw different deductive conclusions from the two sorts of assertion. In contrast, the theories summarized in Table 1 draw no such distinction in the meaning and logic of causes and enabling conditions. Similarly, probabilistic theories have difficulty in distinguishing between causes and enabling conditions, and predict that no necessary conclusions will be drawn from them. Third, the principle of causal deduction implies that some valid deductions should be easier than others: those based on mental models should be drawn more accurately than those that call for fully explicit models. Theories based on schemas and rules might be framed to make the same prediction. Only the model theory, however, predicts the occurrence of "illusory" inferences arising from the failure to represent what is false. In what follows, we describe a series of experiments designed to test these predictions.

3. Studies of causal meanings

The aim of our initial experiment was to test the model theory's prediction that naive individuals should treat some states as possible and some as impossible for each causal relation, and the prediction that *cause* differs in meaning from *allows.* In addition, however, according to the principle of truth, the first true possibility the participants should envisage should correspond to the explicit mental model of the relation. Hence, for *A will cause B,* the first possibility that the participants envisage should be A and B; for *A will prevent B,* the first possibility should be A and not-B; for *A will allow B,* the first possibility should be A and

B; and for *A will allow not B*, the first possibility should be A and not-B. Because it should be quite hard to flesh out the mental models to make them fully explicit, the status of the remaining true possibilities, which correspond to those only in fully explicit models, should be more difficult for the participants to determine, and so they may fail to enumerate them correctly.

3.1. Experiment 1: modal meanings

To understand the meaning of an assertion is at least to grasp what it allows as possible and what it rules out as impossible. Because no previous study had examined this aspect of causal relations, our first experiment was a study of possibilities. The participants' task was to list what was possible and what was impossible given each of the four weak causal assertions. We distinguish this task from one in which participants have to decide whether an assertion is true or false in various situations. This second task calls for a meta-linguistic grasp of the predicates "true" and "false," whereas listing possibilities, according to the model theory, taps into a more fundamental ability to grasp the meaning of assertions (see Johnson-Laird & Barres, 1994).

The assertions in the experiment concerned singular causal relations with unknown outcomes, and so they were of the form:

1. A will cause B.
2. A will prevent B.
3. A will allow B.
4. A will allow not B.

We also included a control assertion, which was a tautology of the following form:

5. A or not A will cause B or not B.

which is compatible with all four possibilities. To ensure that there was no confounding between contents and causal relations, we used five different contents rotated over the five assertions.

Method. The participants listed both possible cases and impossible cases for the 25 assertions, which were presented in a different random order to each participant. The five contents, which were rotated over the relations, concerned health and the performance of mechanical systems. We aimed to select materials that were meaningful to the participants but that made sense in all of the causal relations. The contents are illustrated here:

1. Generating a strong reactivity will cause damage to the reactor.
2. Having a spinal implant will prevent Vivien from being in pain.
3. Using the new fuel will allow the engine to survive the test.
4. Eating protein will allow her not to gain weight.
5. Running the new application will cause the computer to crash.

The participants were presented with an assertion, and their task was to write down what was possible and what was impossible given the assertion. They were free to write the items in any order. And they were told explicitly that there were four cases to consider for each

Table 4
The patterns of response to the five causal relations in Experiment 1, and the number of participants (n = 20) generating each of them on at least three trials out of five. The table shows only those interpretations made by more than one participant.

	The participants' interpretations																				
Listed as possible	a	b	a	b	a	b	a	b							a	b					
	a	¬b					a	¬b	a	¬b			a	¬b	a	¬b					
	¬a	b	¬a	b					¬a	b			¬a	b	¬a	b					
	¬a	¬b	¬a	¬b	¬a	¬b	¬a	¬b	¬a	¬b											
Listed as impossible	—		a	¬b	a	¬b					a	b	a	b							
					¬a	b	¬a	b													
													¬a	¬b	¬a	¬b					Totals
Tautology:	13		6																		19
A will cause B:			9		10																19
A will allow B:	4				10		5														19
A will prevent B:									3				14								17
A will allow not B:													10		7						17

sentence, and they were urged to try to deal with all of them. We tested 20 Princeton undergraduates individually, and they were paid $4 for their participation.

Results and Discussion. Table 4 presents the most frequent interpretations for each of the five sorts of assertion, and the numbers of participants who made each interpretation three or more times out of the five trials. The results corroborated the predictions of the model theory. First, the participants generated both true and false possibilities for the causal relations (19 participants did so more often than not, Binomial $p < .0001$). Second, the participants tended to start their list with the possibility corresponding to the explicit mental model of the causal relation (all 20 participants did so more often than not, $p = .5^{20}$). Third, the participants sometimes failed to list all four of the possibilities as either true or false. They listed a mean number of 3.75 possibilities, and seven of the participants failed to list all four possibilities on at least one occasion. As Table 4 shows, the participants tended to make "strong" interpretations, that is, to minimize the number of possibilities that are true. Overall, their interpretations coincided with the model theory's predictions. There are 16 possible interpretations, and so the chance probability of making a predicted strong or weak interpretation is 1/8. There was a significant tendency to make the predicted interpretations for *cause* (Binomial, $p < 1$ in a billion) and for *prevent* (Binomial, $p < 1$ in a billion). The only unexpected result was the tendency of *allow* and *allow not* to elicit strong interpretations. Nevertheless, there was a reliable tendency to treat *allow* as signifying either all possibilities or the three possibilities taking into account the implicature (Binomial, $p < .0003$), and to treat *allow not* as signifying the predicted interpretation (Binomial, $p < .01$).

If causal relations are probabilistic, then all four cases are possible and nothing is impossible, because it is only the relative frequencies of the different cases that matter. But, contrary to probabilistic theories, the participants generated both true and false possibilities

for each causal relation. Skeptics might suppose that this result reflects the demand characteristics of the experiment, and particularly the instruction to list what was possible and impossible. However, more than half the participants were prepared to list all four possibilities as true in the case of the tautology. Likewise, they showed a reliable consensus about what possibilities were ruled out by each sort of assertion. Hence, the demand characteristics of the experiment cannot explain the results. The results also showed that naïve individuals do distinguish between the meaning of causes and enabling conditions (contrary to the theories in Table 1). We carried out a similar experiment using general causal claims, and its results also corroborated the model theory. In sum, both experiments supported the model theory.

3.2. Experiment 2: causes and enabling conditions

If individuals are given a set of possibilities, it is plausible to suppose that they should be able to produce an appropriate causal description of them. But, previous studies in which participants have to describe sets of possibilities have merely resulted in descriptions of each separate possibility in a long disjunction (see e.g., Byrne & Johnson-Laird, 1992). Cheng & Novick (1991) used an alternative way to give individuals a set of possibilities. They presented descriptions of the circumstances without using any causal expressions. The participants' task was to identify the causes and the enabling conditions within these descriptions. The model theory predicts that they should be able to carry out this task. Moreover, if the theory is correct, then they ought to be able to discern the difference between the two in the same circumstances. Cheng and Novick reported such an effect, but they relied on scenarios with constant enabling conditions and inconstant causes. Our next experiment accordingly examined circumstances in which neither causes nor enabling conditions were constant.

The aim of the experiment was to determine whether naïve individuals could distinguish causes and enabling conditions when neither was constant in the circumstances of the same scenario. Consider, for example, the following description:

1. Given that there is good sunlight, if a certain new fertilizer is used on poor flowers, then they grow remarkably well. However, if there is not good sunlight, poor flowers do not grow well even if the fertilizer is used on them.

Logically speaking, we can paraphrase and abbreviate this description as follows:

If sunlight then if fertilizer then growth; and if not sunlight then not growth.

The corresponding circumstances are the following fully explicit models of the possibilities, as constructed by the computer program:

Sunlight	Fertilizer	Growth
Sunlight	¬Fertilizer	Growth
Sunlight	¬Fertilizer	¬Growth
¬Sunlight	Fertilizer	¬Growth
¬Sunlight	¬Fertilizer	¬Growth

In these circumstances, sunlight is an enabling condition for the flowers to grow well:

Sunlight	Growth
Sunlight	¬Growth
¬Sunlight	¬Growth

But, the presence of sunlight also makes the fertilizer act as a cause of the flowers growing well.

In contrast, consider the following description:

2. Given the use of a certain new fertilizer on poor flowers, if there is good sunlight then the flowers grow remarkably well. However, if the new fertilizer is not used on poor flowers, they do not grow well even if there is good sunlight.

This yields the following fully explicit models of the possibilities:

Sunlight	Fertilizer	Growth
¬Sunlight	Fertilizer	Growth
¬Sunlight	Fertilizer	¬Growth
Sunlight	¬Fertilizer	¬Growth
¬Sunlight	¬Fertilizer	¬Growth

The causal roles have been swapped around: the fertilizer is the enabling condition and its presence makes the sunlight act as the cause of the flowers' growth. In these scenarios, cause and enabling condition differ in their meaning and logic. The experiment accordingly replicated Cheng & Novick's (1991) study but used scenarios in which neither causes nor enabling conditions were constant.

Method. We prepared eight pairs of descriptions, such as the fertilizer examples above, in which there were two precursors to the effect and their respective roles as enabling condition and cause were counterbalanced in the two descriptions. We also prepared versions of the pairs of descriptions in which we reversed the order of mention of cause and enabling condition. Thus, corresponding to the preceding pair of examples, there were descriptions as follows:

1′. If a certain new fertilizer is used on poor flowers, then given that there is good sunlight, they grow remarkably well. However, if there is not good sunlight, poor flowers do not grow well even if the fertilizer is used on them.

2′. If there is good sunlight, then given the use of a certain new fertilizer, poor flowers grow remarkably well. However, if the new fertilizer is not used on poor flowers, they do not grow well even if there is good sunlight.

The participants acted as their own controls and read eight descriptions that each concerned a cause, an enabling condition, and an effect. They also read two filler items–one in which there were two joint causes and one in which there were no causes. Each participant encountered just one version of a particular description, but two instances of the four sorts

of description in the experiment as a whole. The order of presentation was randomized for each participant.

We tested 20 Princeton Undergraduates, who were fulfilling a course requirement. Their task was to identify the enabling condition and the cause in each scenario. As in Cheng and Novick's study, the participants were given a minimal instruction about the difference between causes and enabling conditions. They were told that the cause of an event *brings about the event,* and the enabling condition *makes the event possible,* but these terms did not occur in the scenarios themselves. The participants were told that some passages did not contain an enabling condition.

Results and discussion. The participants correctly identified the enabling conditions and causes on 85% of trials, and every participant was correct more often than not ($p = .5^{20}$). Likewise, all the eight scenarios conformed to this trend ($p = .5^{8}$). Hence, individuals given descriptions of scenarios that make no reference to causation can distinguish enabling conditions from causes. Contrary to a long-established tradition from Mill (1874) onwards, we conclude that causes and enabling conditions *do* differ in meaning, that naive individuals can distinguish between them, and that they can base their distinction on independent descriptions of the relevant sets of possibilities. These descriptions rely essentially on conditionals, and, unlike those of Cheng and Novick, the contrast is not based on making enabling conditions constant in the scenario and causes inconstant. It follows that none of the contrasts in Table 1 is essential to the distinction between causes and enabling conditions. They can at best make predictions about only one of our matched pairs of scenarios in which the causal roles are swapped around. Likewise, as we will show in the General Discussion, the results count against probabilistic theories of causation. Are there cases where prior knowledge blocks the swapping around of cause and effect (Daniel Schwartz, personal communication)? Indeed, there are likely to be such cases. Knowledge modulates the interpretation of assertions and can rule out possibilities compatible with them (see Johnson-Laird & Byrne, 2001, for an account of this process). Thus, consider the following description:

> Given that there is sunlight, if there is water, then the plants grow. However, if there is no sunlight, the plants do not grow even if there is water.

It ought to yield possibilities corresponding to those for description 1 above. However, your knowledge of the need for water is likely to eliminate the possibility:

Sunlight ¬Water Growth

and so you will treat sunlight and water as joint causes of growth.

4. Studies of causal reasoning

Reasoning is based on the representation of assertions. In this section, we report three studies that tested the predictions of the model theory about reasoning from causal assertions. The results of the preceding studies showed that naïve individuals distinguish the difference in meaning between *A will cause B* and *A will allow B.* This difference in meaning, which

is predicted by the model theory, should also be borne out in logical reasoning. If naïve individuals are given the following premises:

Eating protein will cause her to gain weight.

She will eat protein.

the model theory predicts that they should tend to draw the conclusion: She will gain weight. But, if they are given analogous premises based on *allow:*

Eating protein will allow her to gain weight.

She will eat protein.

the theory predicts that they should be less likely to draw the conclusion: She will gain weight. Likewise, the following premises based on *prevent:*

Eating protein will prevent her from gaining weight.

She will eat protein.

should tend to elicit the conclusion: She will not gain weight. But, the analogous premises based on *allow-not* should be less likely to elicit this conclusion. The model theory makes analogous predictions about inferences based on the categorical denial of the proposition that occurs in the consequent of the causal claim. Thus, premises of the form:

A will cause B.

Not-B.

should be more likely to elicit the conclusion: Not-A, than analogous premises based on *allow.* Likewise, premises of the form:

A will prevent B.

B.

should be more likely to elicit the conclusion: Not-A, than analogous premises based on *allow-not.*

The principle of causal deduction (see Section 2) implies that some causal deductions should be easier than others. In particular, those for which mental models suffice should yield a high percentage of correct conclusions. For example, consider a problem of the following form:

A causes B.

B prevents C.

What, if anything, follows?

It should be easy to draw a conclusion of the form: A prevents C, because the conclusion holds in the mental model of the premises. As the computer program shows, the premises:

A causes B: a b

. . .

B prevents C: b ¬c

. . .

yield a single integrated mental model:

a b ¬c

. . .

where each row represents a possibility in the model. This model contains the possibilities corresponding to *A prevents C.* This conclusion is valid, because it holds in the fully explicit model of the premises:

a b ¬c

¬a b ¬c

¬a ¬b c

¬a ¬b ¬c

In contrast, consider the following problem:

A prevents B.

B causes C.

What, if anything, follows?

The mental model of the premises is as follows:

a ¬b

b c

. . .

The possibility containing A does not contain C, and so reasoners should draw the conclusion: A prevents C. The conclusion is wrong, however. The fully explicit model of the two premises is:

a ¬b c

a ¬b ¬c

¬a b c

¬a ¬b c

¬a ¬b ¬c

As this model shows, A does *not* prevent C. In fact, there is no causal relation between them.

4.1. Experiment 3: the logical properties of causes and enabling conditions

Experiment 3 was designed to test the predictions about the difference between inferences based on *cause* and *allow,* and *prevent* and *allow not.* The participants were given pairs of premises using these four causal relations coupled with categorical assertions of all four sorts (the antecedent state, the consequent state, and their respective negations). The four key comparisons for the model theory are shown in Table 5 below. The other eight inferences in

Table 5
The 8 crucial forms of inferential problems in Experiment 3. The table presents the percentages of the participants who drew the responses shown in the table.

	The second premise	
	A	Not-B
The first premise:		
A will cause B	∴ B: 93	∴ Not-A: 93
A will allow B	∴ B: 30	∴ Not-A: 47
	The second premise	
	A	B
The first premise:		
A will prevent B	∴ Not-B: 85	∴ Not-A: 82
A will allow not B	∴ Not-B: 36	∴ Not-A: 36

the set of 16 are, in effect, "filler" items, because their validity depends on whether reasoners make strong interpretations of the causal relations.

Method. The participants acted as their own controls and carried out sixteen problems. Each problem consisted of two premises followed by a question, for example:

Eating protein will cause her to gain weight.

She will eat protein.

Will she gain weight?

Yes. No. Perhaps yes, perhaps no.

The participants were told to circle "yes" if the proposition in the question followed logically from the premises, that is, the proposition must be true given that the premises are true, to circle "no" if the *negation* of the proposition in the question followed logically from the premises, that is, the proposition must be false given that the premises are true, and to circle "Perhaps yes, perhaps no" if neither the proposition in the question nor its negation followed logically from the premises. The problems were composed from the four causal relations paired with the affirmation or denial of the antecedent or consequent proposition. We devised 16 sets of contents, half of which concerned singular events about persons and half of which concerned singular events about mechanical systems. We assigned them twice at random to the 16 forms of inference in order to produce two versions of the materials.

The participants were 129 of the best high school graduates in Italy, with a mean age of approximately 19 years, who were applicants to the Scuola Superiore Sant'Anna of Pisa, a highly selective Italian university.

Results and discussion. Table 5 presents the percentages of predicted results for the crucial problems. In each case, the comparisons were in the direction that the model theory predicted, and each of them was highly significant on a Sign test, that is, considerably less improbable than one in a billion. In particular, given premises of the form: A will cause B,

Table 6
The inferences in Experiment 4, their mental models, the conclusions they predict, and the frequencies with which the participants (N = 20) drew them. Valid conclusions are in capital letters.

Second premise	First premise			
	A causes B	A allows B	A prevents B	Not-A causes B
B causes C	a b c . . . A CAUSES C: 20	a b c a A allows . . . C: 18	a ¬b b c A . . . prevents CL 19	¬a b c . . . Not-A CAUSES C: 20
B allows C	a b c a b . . . A allows C: 19	a b c a b a . . . A ALLOWS C: 19	a ¬b b c b . . . A PREVENTS C: 20	¬a b c ¬a b . . . Not-A allows C: 20
B prevents C	a b ¬c . . . A PREVENTS C: 20	a b ¬c a . . . A allows not-C: 14	a ¬b b ¬c . . . A prevents C: 15	¬a b ¬c . . . Not-A PREVENTS C: 20
Not-B causes C	a b ¬b c . . . A prevents C: 9	a b a ¬b c a ¬b c . . . A ALLOWS Not-C: 12	a ¬b c . . . A CAUSES C: 17	¬a b ¬b c . . . Not-A prevents C: 15

A, reasoners drew the conclusion, *B*, reliably more often than from premises of the form: *A will allow B, A* (85 participants fit the prediction, 2 went against it, and the rest were ties). Given premises of the form: *A will cause B, not-B*, reasoners drew the conclusion, *not-A*, reliably more often than from premises of the form: *A will allow B, not-B* (64 participants fit the prediction, 5 went against it, and the rest were ties). Similarly, given premises of the form: *A will prevent B, A*, reasoners drew the conclusion, *not-B*, reliably more often than from premises of the form: *A will allow not B, A* (70 participants fit the prediction, 6 went against it, and the rest were ties). Given premises of the form: *A will prevent B, B*, reasoners drew the conclusion, *not-A*, reliably more often than from premises of the form: *A will allow not B, B* (64 participants fit the prediction, 6 went against it, and the rest were ties). These results show that the pattern of inferences from *cause* and *allow* are quite distinct, and that the pattern of inferences from *prevent* and *allow-not* are also quite distinct. We conclude that the model theory's account of the difference in meaning between the causal relations is upheld. Causes and enabling conditions differ in meaning and hence in the inferences that they support. The result is incompatible with probabilistic theories – a point to which we return in the General Discussion, and it is also incompatible with the many theories that presuppose that causes and enabling conditions do not differ in meaning (see Table 1).

4.2. Experiment 4: deductions from two causal premises

The goal of Experiment 4 was to test the principle of causal deduction. The participants' task was to draw conclusions in their own words from pairs of premises. The first premise

interrelated two states of affairs using one of the four distinct causal relations (*A causes B, A allows B, A prevents B,* and *Not-A causes B*), and the second premise interrelated two states of affairs, B and C, using one of the same set of causal relations. Table 6 presents the 16 pairs of premises and the predictions of the computer program based on mental models. It predicts that reasoners should draw a conclusion in all 16 cases. Half of these conclusions are valid, but half of them are invalid. Those inferences should be difficult for which a discrepancy occurs between the conclusion based on the mental model of the premises and the conclusion based on the fully explicit model of the premises. Naive individuals should tend to draw the conclusion predicted by the mental model. Neither formal rules nor pragmatic schemas, as they are currently formulated, make any predictions about these deductions.

Method. Each participant carried out in a random order all 16 possible inferences based on the four sorts of premise in the following figure:

A - B.

B - C.

The content of the premises consisted of psychological terms that were familiar to the participants, but not so familiar that they would elicit strong beliefs about the truth or falsity of the premises. We used two random allocations of the contents to the set of 16 problems. A typical problem was as follows:

Obedience allows motivation to increase.

Increased motivation causes eccentricity.

What, if anything, follows?

The instructions made clear that although the contents were sensible, it was not crucial for the participants to know precisely what each term meant. Their task was to state in their own words whatever conclusion, if any, followed from the premises, that is, a conclusion must be true given that the premises were true. We tested individually 20 Princeton undergraduates, who received $4 for participating in the experiment.

Results and discussion. Table 6 presents the numbers of participants drawing the main conclusions to each of the sixteen inferences. The participants tended to draw the predicted conclusions whether they were valid (93% of conclusions) or invalid (80% of conclusions). Each participant drew more predicted than unpredicted conclusions ($p = 0.5^{20}$) and 15 out of the 16 inferences yielded more predicted than unpredicted conclusions (Sign test, $p < .0005$). In sum, the results supported the principle of causal deduction. The one inference for which the model theory's prediction failed was of the form:

A causes B.

Not-B causes C.

The model theory predicts the invalid conclusion: A prevents C, which was drawn by nine of the participants. Eight other participants drew the weaker conclusion: A allows not-C. This conclusion is valid, ignoring the implicature that not-A does not allow not-C.

Could the participants have made their responses under the influence of the particular verbs that occurred in the premises? For example, given premises of the form:

A causes B.

B causes C.

they might have been biased to draw a conclusion with the same verb: *A causes C.* Although we cannot rule out the possibility that such "atmosphere" effects sometimes occur, they cannot explain the results as a whole. For example, given premises of the form:

Not-A causes B.

Not-B causes C.

atmosphere predicts the invalid conclusion: Not-A causes C. The model theory, however, predicts a different invalid conclusion: Not-A prevents C, which the majority of participants (75%) drew. Alleged atmosphere effects in other domains of reasoning are likewise open to alternative explanations based on the use of models in reasoning (Bucciarelli & Johnson-Laird, 1999).

4.3. Experiment 5: illusory inferences about causal relations

The previous experimental results could conceivably be explained by a theory based on formal rules or pragmatic schemas. So, is there any way to strengthen our claim that reasoners rely on mental models? The answer depends on an unexpected prediction. According to the principle of truth, mental models fail to represent what is false, and so certain premises give rise to *illusory* inferences: most people draw one and the same conclusion, which seems obvious, and yet which is wrong. Such illusory inferences occur in deductive reasoning based on sentential connectives (Johnson-Laird & Savary, 1999) and quantifiers (Yang & Johnson-Laird, 2000). They occur in modal reasoning about possibilities (Goldvarg & Johnson-Laird, 2000), in probabilistic reasoning (Johnson-Laird et al., 1999), and in reasoning about consistency (Johnson-Laird et al., 2000). The illusions provide a strong support for the model theory, and they are contrary to current rule theories. These theories use only valid rules of inference (see e.g., Rips, 1994; Braine & O'Brien, 1998), and so they cannot account for a phenomenon in which most people draw one and the same *invalid* conclusion.

Because illusory inferences occur in reasoning about possibilities, they should also occur in causal reasoning. As an example of a potential illusion, consider the following problem based on singular causal assertions:

One of these assertions is true and one of them is false:

Marrying Evelyn will cause Vivien to relax.

Not marrying Evelyn will cause Vivien to relax.

The following assertion is definitely true:

Vivien will marry Evelyn.

Will Vivien relax? Yes/No/It's impossible to know.

The mental models of the initial disjunction of premises are as follows:

Marry Relax

¬Marry Relax

The premise that is definitely true eliminates the second of these models, and so it seems to follow that Vivien will relax. But, the models of the initial disjunction fail to represent what is false, that is, when the first premise is true, the second premise is false, and vice versa. If it is false that marrying Evelyn will cause Vivien to relax, but true that Vivien will marry Evelyn, then Vivien may *not* relax. Hence, the premises do not imply that Vivien will relax. The conclusion is an illusion. The correct response is that it is impossible to know what will happen. The aim of the experiment was to test whether naïve individual succumbed to the illusory inferences.

Method. Each participant carried out four inferences in a random order: the illusory inference above, an analogous illusion based on *prevent*, and two control problems for which the failure to represent falsity should not yield errors (see Table 8). The problems concerned everyday matters, as in the example above, and the contents were allocated twice to the four forms of inferences. We tested 20 Princeton Undergraduates individually, who were fulfilling a course requirement. Their task was to state what conclusion, if any, followed from the premises. The instructions emphasized that in each initial pair of premises one assertion was true and one was false.

Results and discussion. Table 7 presents the percentages of participants making each response. One participant responded "impossible to know" to all the problems, including the controls, and so we rejected his data. The remaining participants tended to succumb to the illusions, making them more often than one would expect by chance (Binomial test, $p < .02$). Likewise, they were correct more often on the control problems than on the illusory inferences ($p = .5^{19}$). Readers may worry that the participants merely overlooked that one of the initial assertions was true and one was false. The instructions made this point as clear as possible, and other studies have deliberately varied the rubrics and yet still the participants erred. Their think-aloud protocols showed that they had not overlooked the nature of the rubrics (Johnson-Laird & Savary, 1999). The results accordingly support the model theory, but they are contrary to theories based on formal rules or schemas. These theories contain only logically impeccable rules, and the only way in which they could yield invalid conclusions is by a mistake in their application. Such mistakes, as Rips (1994, p. 385) has rightly pointed out, should have "diverse sources," and so "a unified account of errors seems extremely unlikely." It strains credulity to imagine that errors of this sort could lead most reasoners to one and the same invalid conclusion. In contrast, the model theory predicts the illusions on the grounds that mental models fail to represent what is false.

Table 7
The two illusions and the two control problems of Experiment 5 (stated in abbreviated form). The table shows the percentages of participants (N = 19) making each response.

Illusions		Control problems	
1. One is true and one is false:		1′. One is true and one is false:	
A will cause B.		A will cause B.	
Not-A will cause B.		Not-A.	
Definitely true:		Definitely true:	
A		A	
∴ B:	68	∴ B:	100
Impossible to know:	32		
2. One is true and one is false:		2′. One is true and one is false:	
A will prevent B		A will prevent B.	
Not-A will prevent B		Not-A.	
Definitely true:		Definitely true:	
A		A	
∴ Not-B	53	∴ Not-B	89
Impossible to know:	47	Impossible to know:	11

5. General discussion

We have advanced a new theory of causality as conceived in everyday life by people with no training in philosophy or logic. The theory gives an account of the meaning of causal relations, their mental representation, and their deductive consequences. It depends on five principles:

1. Truth: People represent propositions by constructing mental models in which each model represents what is true in each possibility compatible with the premises.
2. Temporal constraint: If A has a causal influence on B, then B does not precede A in time.
3. Causal modalities: The meaning of a causal relation between two states of affairs, A and B, concerns what is possible and what is impossible in their co-occurrences. The principle applies to general causal claims (represented in a single model of possible states) and to singular causal claims (represented in a set of models of possibilities).
4. Circumstantial interpretation: Causal interpretation depends on how people conceive the circumstances of states, that is, on the particular states that they consider to be possible, whether real, hypothetical, or counterfactual.
5. Causal deduction: Individuals base their causal deductions on mental models of the premises, inferring whatever conclusion, if any, has the possibilities corresponding to those of the premises.

Does the mere existence of the relevant set of possibilities satisfying the temporal constraint suffice for a causal relation? Or, could there be cases that satisfy our analysis but that are not causal? “Many things,” Cheng (1997, p. 367) writes, “follow one another regularly, yet one does not infer a causal relation between them.” She cites the case of a rooster that crows every day just before sunrise. On our analysis, however, the claim that the rooster’s crowing causes the sun to rise means that the rooster cannot crow at any time of

day, including midnight, without the sun rising shortly thereafter. In a fairy tale, such a magic rooster would indeed be said to cause the sunrise (without presupposing any causal mechanism). Some alternative views about causation are that causes and enabling conditions do not differ in meaning, that causation is a probabilistic notion, that its meaning refers to a mechanism, that deductions about causal relations depend on schemas or rules of inference, and that causal meanings and principles of reasoning differ from one domain to another. If our theory is right, then each of these views, which we discuss in turn, is wrong.

5.1. Causes differ in meaning from enabling conditions

The model theory implies that there are four weak causal relations (*A will cause B, A will prevent B, A will allow B,* and *A will allow not-B*). Experiment 1 showed that individuals distinguish between these relations. They tended to generate the predicted sets of possibilities (and impossibilities) for each relation. From Mill (1874) onwards, however, the consensus has been that causes and enabling conditions do not differ in meaning or logic. This view has in turn led psychologists to search for some other difference between them. They have proposed many putative distinctions–enabling conditions are normal and causes abnormal, enabling conditions are common and causes rare, enabling conditions are constant and causes inconstant, and so on (see Table 1). But, as the model theory predicted, naïve individuals in Experiment 1 distinguished between their meanings. Given an assertion, such as: Using the new fuel will cause the engine to survive the test, all but one of the participants listed as impossible the following situation: Use of the new fuel and the engine does not survive the test. But, given the assertion: Using the new fuel will allow the engine to survive the test, most participants listed as impossible the following situation: Not using the new fuel and the engine survives the test.

The model theory postulates that causal status in complex cases is determined by the set of possibilities as a whole, that is, the circumstances. In one set of circumstances, for example, if there is sunlight, then fertilizer causes flowers to grow. In a contrasting set, the causal roles can be swapped round: if there is fertilizer, sunlight causes the flowers to grow. An ideal way to test the difference would be to ask people to put into their own words the relevant sets of possibilities, but there is no magical way to inject a set of possibilities into someone's head. In Experiment 2, we therefore described them in other terms, that is, without using any causal expressions. The participants' task was to identify which was the cause and which was the enabling condition. They were able to do so in a highly reliable way. The experiment thus replicated Cheng & Novick (1991) but with an important difference. It showed that individuals can distinguish between causes and enabling conditions when neither is constant in the circumstances. Likewise, as the experiment showed, they may judge that a passive agent, for example, sunlight or the presence of oxygen, is the cause, and that an active agent, for example, fertilizer or a spark, is the enabling condition. If there is tendency to judge active agents as causes in everyday life, then it may reflect a pragmatic shortcut rather than a full understanding of the circumstances.

The distinction in meaning between causes and enabling conditions yields different logical consequences. Causal premises such as:

Eating protein will cause her to gain weight.

She will eat protein.

imply the consequent: Therefore, she will gain weight. But, enabling premises such as:

Eating protein will allow her to gain weight.

She will eat protein.

do not imply the consequent. She may, or may not, gain weight. Experiment 3 corroborated these predictions: the two sorts of claim are logically distinct. The factors invoked by psychologists to distinguish between causes and enabling conditions (see Table 1) are neither necessary nor sufficient to explain the difference between their logical properties. The difference in their meanings suffices.

5.2. The meanings of causal relations are not probabilistic

Just as a philosophical view about causation led psychologists to discount a semantic distinction between causes and enabling conditions, so too another philosophical view has led them to hold that the meaning of causal relations in everyday life is probabilistic. Probabilistic theories may be viable in metaphysics and in scientific conceptions of causation, especially since the development of quantum mechanics (see e.g., von Mises, 1957, Sixth lecture). Probabilities may enter into the induction of causal relations from observations (Pearl, 1988; Shafer, 1996; Cheng, 1997; Lober & Shanks, 1999). Likewise, a causal assertion may differ in its probability, that is, some causal assertions are highly probable whereas others are highly improbable. But, are we to suppose that causal relations themselves have *meanings* that are probabilistic?

This view would have astonished Hume (1748/1988), who took a causal claim to apply universally with a constant conjunction of cause and effect. It would have astonished Kant, who argued that causation was an a priori notion which demands "that something, A, should be of such a nature, that something else, B, should follow from it necessarily" (Kant, 1781/1934, p. 90). It would have astonished Mill, who wrote: "The invariable antecedent is termed the cause; the invariable consequent, the effect" (Mill, 1874, p. 237). Indeed, Russell (1912–13) took the view that the concept of causation should be expurgated from philosophy, because it had been replaced by probabilistic correlations in science. That the concept of cause itself might be probabilistic appears to be a doctrine unique to the world post quantum mechanics.

The principal evidence lending support to the probabilistic doctrine is that people assent to causal claims even when they know there are exceptions to them. Loose generalizations are endemic in daily life. Hence, most people assent to the proposition:

Smoking causes lung cancer

even though they know that not everyone who smokes gets the disease. But, most people are also likely to assent to the more accurate proposition:

Smoking often causes lung cancer.

Readers who agree that this assertion is more accurate have conceded the main point: if causes were intrinsically probabilistic, then the two assertions would not differ in meaning. Another factor in the use of loose causal generalizations may be that people are well aware that many causes in everyday life yield their effects only if the required enabling conditions are present and the potentially disabling conditions are absent. Naive individuals, as Cummins (1995, 1998) has shown, are sensitive to these factors. Hence, when people assent to the loose generalization about smoking, they are granting the effect other things being equal. As Cummins remarks, they mean that the causal relation holds unless some disabling condition is present.

In our view, a probabilistic *meaning* of causality does not accord with everyday usage, as several strands of evidence show. The first strand of evidence is that naïve individuals in Experiment 1 judged that certain possibilities are ruled out by causal assertions. Given, for example, the assertion: Running the new application will cause the computer to crash, they listed as impossible the following situation: Running the new application and the computer does not crash. Suppose on the contrary that the probabilistic theory were correct. It would follow that the causal assertion means merely that the probability of the computer crashing given that one runs the new application is higher than the probability of it crashing given that one does not run the new application. In this case, it is possible that one runs the new application and the computer does not crash. The only way to refute the causal claim would be to make a set of observations to show that the relative frequencies of crashes supported the difference between the two conditional probabilities. Defenders of the probabilistic approach might counter that the "demand characteristics" of Experiment 1 forced the participants to list both what is possible and what is impossible. But, this claim is refuted by the fact that the majority of the participants treated the tautology as consistent with all possibilities, that is, they did not list any situation as impossible in this case. Jonathan Evans (personal communication) has suggested that the use of singular events rather than general causal claims might have discouraged probabilistic interpretations in Experiment 1. In fact, we have run a study with general claims that yielded very similar results. Moreover, if causation is a probabilistic notion, then it should be probabilistic for both singular and general claims. Consider the following problem:

Smoking causes lung cancer.

Pat smoked but did not get lung cancer.

Why was that?

As we observed anecdotally, people tend to make the following sorts of response: Perhaps Pat had a strong resistance to cancer, he might not have smoked much, or he may have given up smoking. They do not respond: Because that is the nature of causation. Yet, this answer is correct according to the probabilistic theory. In short, Hume (1739/1978) may have been right when he argued that people treat chance as a case of a hidden cause.

The second strand of evidence derives from the distinction between causes and enabling conditions. Consider a set of possibilities stated with their frequencies of occurrence (out of 100), as in the following partition:

Sunlight	Fertilizer	Growth	20
Sunlight	¬Fertilizer	Growth	20
Sunlight	¬Fertilizer	¬Growth	20
¬Sunlight	Fertilizer	¬Growth	20
¬Sunlight	¬Fertilizer	¬Growth	20

It follows that:

p(Growth | Sunlight) > p(Growth | ¬Sunlight), i.e., .66 > 0

p(Growth | Fertilizer) > p(Growth | ¬Fertilizer), i.e., .5 > .33

Hence, both sunlight and fertilizer are causes according to the probabilistic theory, though sunlight has the greater probabilistic contrast. Yet, as the model theory predicts, individuals in Experiment 2 judged sunlight to be the enabling condition and fertilizer to be the cause. The distinction between causes and enabling conditions might be reconstructed within a probabilistic theory in the following way: *A causes B* means that B is very probable given A, and *A allows B* means that B is quite probable given A (Jonathan Evans, personal communication). This idea seems plausible, but it makes exactly the wrong predictions about the example above. In fact, the probabilistic theory obliterates the distinction between causes and enabling conditions. Yet people are sensitive to this distinction, and so the probabilistic theory fails to account for the everyday meaning of causal relations.

A third strand of evidence demonstrates another difficulty for the probabilistic theory. When probabilities are held constant, a manipulation of content can yield different attributions of causality (Legrenzi & Sonino, 1994; White, 1995; Koslowski, 1996). Such results are clearly inexplicable if causation is equivalent to an assertion about conditional probabilities.

A fourth stand of evidence comes from Experiment 3. It showed that naïve individuals tend to draw definite conclusions from causal assertions. Given a causal premise, such as: Eating protein will cause her to gain weight, and the assertion that she eats protein, the majority of participants concluded that she *will* gain weight. But, if a causal assertion is merely a statement of a high conditional probability, then the participants should have refused to draw any deductive conclusion. At the very least, they should have qualified their answers in terms of probabilities.

A fifth strand of evidence is that the probabilistic theory cannot explain cases in which a cause decreases the probability of an effect (see e.g., Salmon, 1980). Here is an example from Tooley (1987, p. 234–5). Disease A causes death with a probability of 0.1, and disease B causes death with a probability of 0.8. Each disease, however, confers complete immunity to the other disease. Given that an individual is in a certain condition, he is bound to contract either disease A or disease B. In fact, a particular individual in this condition contracted disease A, and as a result died. If he hadn't contracted disease A, then he would have contracted disease B, and so his probability of dying would have been 0.8. Hence, the cause of his death (disease A), in fact, did not increase the probability of his death, but lowered it.

A final strand of evidence comes from a paraphrase. A claim such as: If there were no

gravity then planets would not orbit stars, is unequivocal. Yet, it can be paraphrased as: Gravity causes planets to orbit stars. Nothing is uncertain in the meaning of such causal claims. Indeed, to allow some uncertainty, one asserts: Gravity probably causes planets to orbit stars.

5.3. Do causal meanings refer to mechanisms?

At the core of the model theory is the modal principle that causes concern sets of possibilities governed by a temporal constraint. But what of the other metaphysical principle of causal powers or mechanisms? As we mentioned in Section 2, everyday causal assertions often concern "action at a distance," and they sometimes deliberately deny the existence of an underlying causal mechanism. We therefore argued that it may be mistaken to build these metaphysical principles into the everyday meaning of causality. Such a claim, however, does not deny the existence of causal mechanisms.

You understand causal relations better when you have access to a mechanism, that is, you have a dynamic mental model that can unfold over time (Johnson-Laird, 1983, Ch. 15). It allows you to simulate the sequence of events and to infer causal consequences: A will cause B, but B in turn may feed back and have a causal influence on A (Green, 2001). A mechanism also provides a lower level of explanation. At one level, you know what the system does, and at one level down you know how it does it (see Miyake, 1986). Suppose, for example, that you can put your computer into sleep mode in different ways, such as pressing control+A or control+B, or control+A+B. It seems that the two keys A and B have equivalent roles given that you press control: Pressing key A or key B, or both, causes the computer to go to sleep. But, consider the mechanism at one level down, namely, the circuit. If all you can observe are the positions of the keys and their effect on the computer, it is impossible to identify the circuit. In principle, there are infinitely many distinct circuits in which the keys might have their effects, just as there are infinitely many distinct ways in which to compute any computable function. If you can at least measure whether or not current flows through a key, you may be able formulate a description of the mechanism. You might discover, for instance, that when key B is pressed, current flows through it and puts the computer to sleep independently from the position of key A. But, when key B is *not* pressed it completes a series with key A, and so the computer goes to sleep only when A is pressed too. Hence, the only effect of switch A is to close or to break this second circuit. This mechanism at a lower level justifies the description: Pressing key B, or pressing key A when key B is not pressed, puts the computer to sleep.

What causes the current to flow in these various circuits? A causal mechanism itself may be further decomposed into a lower level mechanism, and at each level reference is made to further causes. Hence, we can talk of a difference in potential that causes a flow of electrons in each of the circuits. But, this mechanism itself contains a causal claim: the potential difference causes the flow. Once again, we can ask in turn for its causal mechanism. And so on ad infinitum. Each mechanism at a lower level embodies a causal claim, and, like a child's series of "why" questions, we can go on asking for the mechanism underlying each answer until we run out of knowledge.

The morals are threefold. First, mechanisms are not part of the *meaning* of causal assertions, because causal assertions can deny their existence without contradiction. Second,

a causal mechanism itself embodies causal relations at a lower level than those of the phenomena for which it provides an account. To stipulate that a causal claim is an assertion that a mechanism exists runs the risk of a vicious circle: you appeal to a set of causal relations in order to give an account of the meaning of causation. Third, mechanisms and simulations based on their mental models provide a powerful way of inferring causal consequences. They can be crucial in inferring causation from correlations or other inductive observations (see e.g., White, 1995).

5.4. Causal deduction depends on models, not rules or schemas

Although some theorists argue that causal relations cannot be deduced, but only induced, we have shown that naïve individuals do make causal deductions. A common sort of causal deduction occurs when you use your background knowledge to infer a causal interpretation of a state. You know that an insulin injection prevents a coma in diabetes; your diabetic friend gives herself such an injection; and you infer that she will not go into a coma. Of course, your conclusion may be mistaken: all deductions based on empirical premises, whether causal or not, are defeasible. That is, they may have conclusions that turn out to be false because a premise is false. Likewise, the deductive process is not always straightforward. Santamaria, Garcia-Madruga, and Johnson-Laird (1998) gave participants a pair of believable conditional premises, which each expressed a causal relation, such as:

If Marta is hungry, then she takes an afternoon snack.
If Marta takes an afternoon snack, then she has a light dinner.

Both premises are believable, and they yield the following valid conclusion: If Marta is hungry, then she has a light dinner. Yet, the participants were reluctant to draw this conclusion, presumably because it is unbelievable. It lacks the causal link (eating the snack) between its antecedent and consequent, and so it violates the normal relation between hunger and dinner.

Causal deductions could depend on formal rules of inference or on pragmatic reasoning schemas. Experiment 4, however, corroborated the principle of causal deduction–that naive individuals deduce causal relations from mental models of the premises. The experiment showed that they tended to draw those conclusions supported by mental models of the premises, whether or not the conclusions were valid. Experiment 5 established the occurrence of illusory inferences based on causal premises. Given premises of the form:

One of the following assertions is true and one of them is false:
A will cause B.
Not-A will cause B.
This assertion is definitely true: A

most participants inferred invalidly: B will occur. These illusions are a crucial test of the model theory. It predicts their occurrence because mental models represent only what is true (the principle of truth). Theories based on formal rules or pragmatic schemas, however, cannot account for the illusions. These theories postulate only valid rules of inference, and so they have no way to explain the systematic occurrence of invalid inferences.

5.5. Causal meanings and principles of reasoning do not differ from one domain to another.

Throughout this article, we have presupposed a uniform account of causation and causal reasoning. A contrasting view is that causal reasoning differs from one domain to another. Similarly, individuals may rely on, say, pragmatic reasoning schemas when they have knowledge of a domain, but on mental models or some other "weak" method when they have no relevant knowledge. In fact, a study of how people explain inconsistencies suggests that there are strong uniformities across domains (see Legrenzi et al., 2001). In two experiments, the participants were presented with sets of inconsistent assertions, and their task was to rank order the probability of seven different putative explanations of each inconsistency. The participants tackled four scenarios from five different domains: physics, mechanics, biology, psychology, and socio-economics. They all rated an explanation consisting of a cause and an effect as more probable than either the cause alone or the effect alone. This bias was uniform over all five domains. Likewise, in the present studies, no striking effects occurred as a result of the content of the problems, for example, health problems versus mechanical problems in Experiment 1. There may be differences yet to be detected in causal reasoning depending on domain or knowledge. So far, however, the hypothesis of uniform principles has survived testing.

5.6. Induction and the model theory

The model theory has implications for the process of inducing causal relations from observations. Knowledge of explicit possibilities, as we have seen, allows individuals to deduce an appropriate causal relation from observations. It can also override data about the relative frequencies of different sorts of event (see Section 2). But the theory of meaning has implications for the interpretation of frequency data even where no knowledge is available to aid the process. According to the probabilistic account of cause, the appropriate strategy for interpreting a probability distribution is to assess the difference between the two conditional probabilities, $p(B|A)$ and $p(B|\neg A)$, where A is the putative cause and B is the effect. If the distribution is in the form of a 2×2 table of frequencies, then the data in each cell have to be taken into account to compute these conditional probabilities. Naive individuals sometimes carry out this strategy (Lober & Shanks, 1999), but often they fail to do so (Beyth-Marom, 1982; Mandel & Lehman, 1998). Some theorists have defended the failure as adaptive (cf. Anderson & Sheu, 1995; Gigerenzer & Hoffrage, 1995; 1999), but Over & Green (2001) argue convincingly that the correct Bayesian response must take into account all four cells in the table. However, if causation has the meaning shown in Table 2, then it is not necessary to take into account all four frequencies in order to establish a causal relation. Indeed, what matters is not how often something occurs, but whether or not it occurs (see Schwartz et al., 1998).

Suppose you want to test whether or not a set of observations supports the following causal relation: Heating water to 100 °C in normal air pressure causes it to boil. If you observe cases of boiling water at 100 °C, and no cases of water at 100 °C failing to boil, then you are almost there. It suffices to establish, if you don't already know, that water is not

invariably boiling. The frequencies of the various observations do not matter. Whether or not water boils at another temperature is not strictly relevant to the causal claim. If it doesn't, then heating it to 100 °C is the unique cause of boiling; if it does, then heating it to 100 °C is merely one cause of boiling. In more abstract terms, a causal relation of the form *A causes B*, can be inferred from the occurrence of cases of A and B, and the absence of cases of A and not-B, granted that B is not invariable. The *mental* models of the relation make explicit, however, only the conjunction of A and B. Hence, the theory predicts that there should be a bias towards this contingency in assessments of causation. Such a bias occurs, particularly in situations in which there is more than one putative cause (Schustack & Sternberg, 1981). Otherwise, individuals do tend to take into account cases of A and not-B (Einhorn & Hogarth, 1978). With binary variables in the form of "present" versus "absent," they are likely to focus on the co-occurrence of the cause and effect. But, if both values of a variable have to be represented in models, for example, "high" versus "low," then participants are more likely to take into account falsifying cases (Beyth-Marom, 1982).

When cases of A and not-B *do* occur, there are two possibilities: either A does not cause B, or A may have occurred in the absence of an enabling condition – or, equivalently, in the presence of a disabling condition. It becomes necessary to make observations of the co-occurrence of potential enablers and disablers, and the circumstances may be as complex as those that we spelled out for Experiment 2. It is not easy to determine the correct relations. An effect may have multiple alternative causes; it may have multiple enabling conditions (Hart & Honoré, 1959, 1985). Where there are competing causes, as the model theory predicts, people tend to focus on a single cause and to discount other potential causes (Shaklee & Fischhoff, 1982), just as they focus on single options in decision making (Legrenzi et al., 1993) and single explanations in induction (Sloman, 1994).

6. Conclusion

Our results substantiate the model theory of causal relations. The theory is founded on a few simple, but powerful, principles. The meanings of causal relations are sets of possibilities in which an effect cannot precede a cause. Naive individuals can envisage these possibilities in fully explicit models, but with complex descriptions they tend to rely on mental models representing only what is true in the possibilities. They deduce the consequences of causal claims from what holds in their mental models of the premises. Deductions that they cannot make in this way are likely to be beyond them.

Acknowledgments

We thank Paolo Legrenzi and Maria Sonino Legrenzi for carrying out Experiment 3, Riccardo Varaldo, Director of the Scuola Superiore Sant'Anna of Pisa, for allowing us to test applicants to the school, and Stefania Pizzini for help in analyzing the results. We thank Ruth Byrne, Denise Cummins, Jonathan Evans, Vittorio Girotto, James Greeno, Daniel Schwartz, Steve Sloman, and Paul Thagard, for their constructive comments on an earlier version of

this paper. We are grateful to many colleagues for their help: Patricia Barres, Victoria Bell, Zachary Estes, David Green, Denis Hilton, Patrick Lemaire, Bradley Monton, Hansjoerg Neth, Mary Newsome, David Over, Sergio Moreno Rios, David Shanks, Larry Solan, Vladimir Sloutsky, Jean-Baptiste van der Henst, and Yingrui Yang. Part of the research was presented to the Workshop on Deductive Reasoning at London Guildhall University in January 1999, and the authors are grateful to its organizer, David Hardman, and the participants for their advice. The research was supported in part by NSF Grant 0076287.

References

Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanism: an explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31,* 82–123.

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54,* 299–352.

Anderson, J. R., & Sheu, C-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition, 23,* 510–524.

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22,* 25–51.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory & Cognition, 10,* 511–519.

Braine, M. D. S., & O'Brien, D.P. (Eds.). (1998). *Mental Logic.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247–303.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In Friedman, W.J. (Ed.) *The Developmental Psychology of Time.* pp. 209–254. Orlando, FL: Academic Press.

Burks, A. (1951). The logic of causal propositions. *Mind,* LX, 363–382.

Byrne, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In Medin, D.K. (Ed.) *The psychology of learning and motivation: advances in research and theory,* Vol. 37 (pp. 105–154). San Diego, CA: Academic Press.

Byrne, R. M. J., & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564–575.

Byrne, R. M. J., & Johnson-Laird, P.N. (1992). The spontaneous use of propositional connectives. *Quarterly Journal of Experimental Psychology, 45,* 89–110.

Carnap, R. (1966). *Philosophical Foundations of Physics.* New York: Basic Books.

Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review, 104,* 367–405.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328.

Cheng, P. W., & Nisbett, R. E. (1993). Pragmatic constraints on causal deduction. In Nisbett, R.E. (Ed.) *Rules for reasoning.* (pp. 207–227). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58,* 545–567.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40,* 83–120.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: an introduction to psycholinguistics.* New York: Harcourt Brace Jovanovich.

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition, 23,* 646–658.

Cummins, D. D. (1998). The pragmatics of causal inference. *Proceedings of the twentieth annual conference of the cognitive science society* (p. 9).

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

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99,* 3–19.

Fisher, R. A. (1959). *Smoking.* London: Oliver & Boyd.

Garnham, A., & Oakhill, J. V. (1996). The mental models theory of language comprehension. In Britton, B.K., & Graesser, A.C. (Eds.) *Models of understanding text.* (pp. 313–39). Hillsdale, NJ: Erlbaum.

Geminiani, G. C., Carassa, A., & Bara, B. G. (1996). Causality by contact. In Oakhill, J., & Garnham, A. (Eds.) *Mental models in cognitive science.* (pp. 275–303). Hove, East Sussex: Psychology Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review, 102,* 684–704.

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: a reply to Lewis & Keren (1999) and Mellers & McGraw (1999). *Psychological Review, 106,* 425–430.

Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica, 78,* 111–133.

Goldvarg, Y., & Johnson-Laird, P. N. (2000). Illusions in modal reasoning. *Memory & Cognition, 28,* 282–294.

Green, D. W. (2001). Understanding microworlds. *Quarterly Journal of Experimental Psychology,* in press.

Green, D. W., & Over, D. E. (2000). Decision theoretic effects in testing a causal conditional. *Current Psychology of Cognition, 19,* 51–68.

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

Harré, R., & Madden, E. H. (1975). *Causal powers.* Oxford: Blackwell.

Hart, H. L. A., & Honoré, A. M. (1985). *Causation in the law.* (2nd ed.) Oxford: Clarendon Press. (First edition published in 1959.).

Hesslow, G. (1988). The problem of causal selection. In Hilton, D.J. (Ed.) *Contemporary science and natural explanation: commonsense conceptions of causality* (pp. 11–32). Brighton, Sussex: Harvester Press.

Hilton, D. J., & Erb, H-P. (1996). Mental models and causal explanation: judgements of probable cause and explanatory relevance. *Thinking & Reasoning, 2,* 273–308.

Hume, D. (1988). *An Enquiry Concerning Human Understanding.* A. Flew. (Ed.) La Salle, IL: Open Court. (Originally published 1748.)

Hume, D. (1978). *A Treatise on Human Nature.* L.A. Selby-Bigge (2nd ed.). Second ed. Oxford: Oxford University Press. (Originally published 1739.)

Johnson-Laird, P. N. (1983). *Mental models: towards a cognitive science of language, inference and consciousness.* Cambridge: Cambridge University Press; Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., & Barres, P. E. (1994). When 'or' means 'and': a study in mental models. *Proceedings of the sixteenth annual conference of the cognitive science society,* 475–478.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. M. J. (2001). Conditionals: a theory of their meaning, pragmatics, and role in inference. Under submission.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288,* 531–532.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., Legrenzi, M. S., & Caverni, J-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88.

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: a novel class of erroneous deductions. *Cognition, 71,* 191–229.

Kahneman, D., & Miller, D. T. (1986). Norm theory: comparing reality to its alternative. *Psychological Review, 93,* 75–88.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.) *Judgment under uncertainty: heuristics and biases.* Cambridge: Cambridge University Press.

Kant, I. (1934). *Critique of pure reason.* Translated Meiklejohn, J.M.D. New York: Dutton. (Originally published 1781.)

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28,* 107–128.

Koslowski, B. (1996). *Theory and evidence: the development of scientific reasoning.* Cambridge, MA: MIT Press.
Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition, 49,* 37–66.
Legrenzi, M., Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (2001). Reasoning to consistency: a theory of naïve nonmonotonic reasoning. Under submission.
Legrenzi, P., & Sonino, M. (1994). Psychologistic aspects of Suppes's definition of causality. In Humphreys, P. (Ed.) *Patrick Suppes: scientific philosopher, Vol, 1* (pp. 381–399). The Netherlands: Kluwer.
Leslie, A. M. (1984). Spatiotemporal contiguity and perception of causality in infants. *Perception, 13,* 287–305.
Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM [Theory of Mind Mechanism]. *Cognition, 50,* 211–238.
Lewis, C. (1986). A model of mental model construction. *Proceedings of CHI '86 conference on human factors in computer systems.* New York: Association for Computing Machinery.
Lewis, D. (1973). *Counterfactuals.* Oxford: Blackwell.
Lober, K., & Shanks, D. R. (1999). Experimental falsification of Cheng's (1997) Power PC theory of causal induction. *Psychological Review,* in press.
Lombard, L. M. (1990). Causes, enablers, and the counterfactual analysis. *Philosophical Studies, 59,* 195–211.
Mackie, J. L. (1980). *The cement of the universe: a study in causation.* (2nd ed.) Oxford: Oxford University Press.
Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgements of cause, covariation and probability. *Journal of Experimental Psychology: General, 127,* 269–285.
Markovits, H., & Savary, F. (1992). Pragmatic schemas and the selection task: to reason or not to reason. *Quarterly Journal of Experimental Psychology,* 45A, 133–148.
McArthur, L. (1972). The how and what of why: some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology, 22,* 171–193.
McEleney, A., & Byrne, R. M. J. (2001). Counterfactual thinking and causal explanation. Under submission.
McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology, 57,* 189–200.
Michotte, A. (1963). *The perception of causality.* London: Methuen. (Originally published 1946.)
Mill, J.S. (1874) *A system of logic, ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific evidence.* (8th ed.) New York: Harper. (First edition published 1843.)
Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge, MA: Harvard University Press.
Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10,* 151–177.
Morris, M. W., & Nisbett, R. E. (1993). Tools of the trade: deductive schemas taught in psychology and philosophy. Nisbett, R.E. (Ed.) *Rules for reasoning* (pp. 228–256). Hillsdale, NJ: Lawrence Erlbaum Associates.
Newstead, S. E., Ellis, M. C., Evans, J.St.B.T., & Dennis, I. (1997). Conditional reasoning with realistic material. *Thinking & Reasoning, 3,* 49–76.
Osherson, D. N. (1974–6). *Logical abilities in children, Vols. 1–4.* Hillsdale, NJ: Lawrence Erlbaum Associates.
Over, D. E., & Green, D. W. (2001). Contingency, causation, and adaptive inference. *Psychological Review,* in press.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Revised Second Printing. San Francisco: Morgan Kaufmann.
Reichenbach, H. (1956). *The direction of time.* Berkeley: University of California Press.
Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.
Russell, B. A. W. (1912–13). On the notion of cause. *Proceedings of the Aristotelian Society, 13,* 1–26.
Salmon, W. C. (1980). Probabilistic causality. *Pacific Philosophical Quarterly, 61,* 50–74.
Santamaria, C., Garcia-Madruga, J. A., & Johnson-Laird, P. N. (1998). Reasoning from double conditionals: the effects of logical structure and believability. *Thinking and Reasoning, 4,* 97–122.
Schaeken, W. S., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60,* 205–234.

Schustack, M. W. (1988). Thinking about causality. In Sternberg, R.J. & Smith, E.E. (Eds.) *The Psychology of thinking* (pp. 92–115). Cambridge: Cambridge University Press.

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

Schwartz, D. L., Goldman, S. R., Vye, N. J., Barron, B. J., & CTGV. (1998). Aligning everyday and mathematical reasoning: the case of sampling assumptions. In Lajoie, S. (Ed.) *Reflections on statistics: agendas for learning, teaching and assessment in K-12.* (pp. 233–274). Mahwah, NJ: Erlbaum.

Shafer, G. (1996). *The art of causal conjecture.* Cambridge: MIT Press.

Shaklee, H., & Fischhoff, B. (1982). Strategies of information search in causal analysis. *Memory & Cognition, 10,* 520–530.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47,* 1–51.

Sloman, S. A. (1994). When explanations compete: the role of explanatory coherence on judgments of likelihood. *Cognition, 52,* 1–21.

Stalnaker, R. (1968). A theory of conditionals. In Rescher, N. (Ed.) *Studies in logical theory.* pp. 98–112. Oxford: Blackwell.

Stevenson, R. J. (1993). *Language, thought and representation.* New York: Wiley.

Suppes, P. (1970). *A probabilistic theory of causality.* Amsterdam: North-Holland.

Suppes, P. (1984). *Probabilistic metaphysics.* Oxford: Basil Blackwell.

Taylor, R. (1966). *Action and purpose.* Englewood Cliffs, NJ: Prentice-Hall.

Thompson, V. A. (1995). Conditional reasoning: the necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49,* 1–60.

Tooley, M. (1987). *Causality: a realist approach.* Oxford: Oxford University Press.

Turnbull, W., & Slugoski, B. R. (1988). Conversational and linguistic processes in causal attribution. In Hilton, D. (Ed.) *Contemporary science and natural explanation*: *commonsense conceptions of causality* (pp. 66–93). Brighton, Sussex: Harvester Press.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In Fishbein, M. (Ed.) *Progress in social psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Vendler, Z. (1967). *Linguistics in philosophy.* Ithaca, NY: Cornell University Press.

von Mises, R. (1957). *Probability, statistics and truth.* Second revised English edition based on the third German Edition of 1951. London: Allen & Unwin.

von Wright, G. H. (1973). On the logic and epistemology of the causal relation. In Suppes, P. (Ed.) *Logic, methodology and philosophy of science, IV* (pp. 293–312). Amsterdam: North-Holland.

Wason, P. C. (1966). Reasoning. In Foss, B.M. (Ed.) *New horizons in psychology.* Harmondsworth, Middlesex: Penguin.

White, P. A. (1995). Use of prior beliefs in the assignment of causal roles: causal powers versus regularity-based accounts. *Memory and Cognition, 23,* 243–254.

Yang, Y., & Johnson-Laird, P. N. (2000). Illusions in quantified reasoning: how to make the impossible seem possible, and *vice versa. Memory & Cognition, 28,* 452–465.