

The Role of Mechanism Beliefs in Causal Reasoning

Woo-kyoung Ahn

Yale University

Charles W. Kalish

University of Wisconsin, Madison

To appear in F. Keil (Ed.) Explanation and Cognition, MIT press

(14-Jan-02)

## **Introduction: Characterizing the Questions of causal reasoning**

This chapter describes the mechanism approach to the study of causal reasoning. We will first offer a characterization of the central issues in human causal reasoning, and will discuss how the mechanism approach addresses these issues. In the course of this presentation, we will frequently compare the mechanism approach with alternative accounts based on analyses of covariation, or what is often termed the regularity view. The aims of this chapter are the following: to explain why covariation and mechanism are different, to discuss why such a distinction is actually a useful tool for our understanding of causal reasoning, and to explicate the complementary nature of the two views.

Before presenting these two approaches, it is necessary first to offer a description of the domain or problem itself : namely, what are these alternative approaches to? Although there are a number of different ways of characterizing the study of causal reasoning, we depict it as an attempt to examine how people think about causal relations and how people identify causes. That is, our focus is descriptive. Hence, in this chapter, we do not try to provide a normative account of how people should think about causes, nor do we take as our goal a description of how people are able to reason under exceptional circumstances. Glymour (1998) has argued that proponents of a mechanism view commit an "ecological fallacy" by concentrating on what people do most often, most typically, or perhaps, most naturally. We would amend Glymour's characterization slightly and argue that ecological validity is among the virtues of a focus on mechanism. Adopting a descriptive perspective is a reasonable and valuable approach for psychologists interested in characterizing how people actually carry out the task of causal reasoning.

There are two parts to the question of how people identify causal relations. First, what do people mean when they identify a relation as causal? Second, by what process do people identify a relation as causal? These two questions roughly correspond to the distinction made by Rey (1983) with regard to categorization. On the one hand, argues Rey, there are metaphysical questions of categorization: What do people think makes something a member of one category rather than another? Given our focus on causation, we may ask what people think distinguishes a causal

relation from a non-causal one. Hence, this question has to do with the *definition* of cause. In addition to metaphysical questions, Rey points out that we are often interested in epistemic questions: What do people do to decide whether or not something is a member of a category? Again, with respect to causation, we may ask what people do when they are trying to decide if a relation is causal or not, or to determine a cause in a given situation. Thus, this question has to do with *methods of identifying* causes. Although these two questions are clearly linked, an answer to one does not necessarily provide an answer to the other. There may be aspects of a definition that are not (typically) used in identification, and means of identifying instances that are only tenuously connected to definitions.

Most work in causal reasoning has focused on epistemic questions (e.g., how people identify a relation as causal). The mechanism approach is most centrally a claim about people's definitions or conceptions of causality. However, one claim of this approach is that people's strategies for identifying instances of causal relations typically derive from their beliefs about the nature of causation.

In presenting the mechanism view, we will first describe its claims about people's conceptions of causation. From this description we will go on to derive hypotheses regarding the process of causal identification. In describing the definitional and process aspects of the mechanism approach, it will often be useful to contrast our view with covariation approaches in order to highlight important characteristics of the mechanism approach. Given recent accounts of conflict between covariation and mechanism approaches, we will conclude by considering some of the relations between covariation and mechanism in causal reasoning.

## **Definition of Causation**

### **Mechanism View**

We believe that the core component of the idea of “cause” is a sense of force. If A causes B, then A makes B happen: B had to happen given A. It was no accident. It is this sense of necessity that distinguishes "real" causal relations from "mere" correlations.

While this characterization is probably universally accepted, the mechanism approach suggests that there is usually more to people's notions of cause. We argue that when people say A causes B, they believe that there is a process that took place between A and B in which a force or causal power is transmitted. Borrowing from Salmon's (1984) example, imagine a spotlight placed in the middle of the ceiling of a dome. The spotlight is a cause for the light projected on the wall because a pulse of light travels from the spotlight to the wall. Now imagine that the spotlight rotates so that the light moves around the wall. Even though there is a regular and reliable succession such that the position of the light at time T1 is always followed by a new position at time T2, we do not say that the position at T1 causes the position at T2 because we know that causal power is not transmitted from one spot of the wall to the other spot of the wall. Likewise, we argue that people's conception of cause includes a causal process that transmits causal power or carries a mark (see also Harré & Madden, 1975; Shultz, 1982 for a similar proposal).

The first claim of the mechanism approach is thus that people believe that there is a more basic process underlying the cause-effect relation. In other words, when people conceive of cause-effect relations, there is a basic assumption of mechanism; the assumption that underlying two causally linked events, there is a system of connected parts that operate or interact to make or force an outcome to occur. Consider getting sneezed on and getting sick. If people think the sneeze is the cause, then they also believe that there must have been a basic process or mechanism by which the sneeze forced the illness to come about. In modern western cultures, we typically understand the mechanism to be infection; getting sneezed on infects you with germs that make you sick. A relatively elaborated notion of the mechanism might include the ideas that germs possess the causal power of making a person sick, that the person's immune system has causal power to counteract germs, and that the person's immune system can be weakened by lack of sleep<sup>1</sup>. This way,

---

<sup>1</sup> As in this example, a mechanism usually consists of multiple factors, each possessing causal power. Of these many interconnected parts in the mechanism, pragmatic constraints would determine which one is "the" cause and which ones are preconditions. These pragmatic constraints would include conversational maxims (Hilton, 1995), contrast / focal sets (Cheng & Novick, 1991; McGill, 1989), abnormality of the factors (Hilton & Slugoski, 1986; Mackie, 1974), and so on.

individuals have more or less elaborated beliefs about the mechanisms mediating between causes and effects.

The second claim of the mechanism view is that mechanisms are framed at a different level of analysis than are the cause and the effect. That is, mechanisms involve theoretical constructs, constructs that are removed from and underlying the evidential phenomena themselves (Gopnik & Wellman, 1994). Thus in the example above, "germ" is part of a theoretical vocabulary that is described at a level different from "sneezed on" and "sick".

Of course this leads to a problem of potentially infinite regress. For instance, in explaining why John had a traffic accident, one might refer to a mechanism of drunk driving. In explaining why drinking causes a traffic accident, one can further provide an explanation beyond this level of the event description (e.g., a person's reaction time is slower when drunk, and so on). Ultimately, this process would be bottomed out to the point where objects must have their causal powers "essentially" with no other more basic process responsible (Harré, 1988). When the issue of essential properties does arise, it is typically dealt with in a non-intuitive, theoretical manner. For instance, accounts of the ultimate or final level of causal relation, beyond which we can no longer pursue questions of mechanism, are typically couched in highly theoretical terms (e.g., quantum theory) or exist within the province of religious explanations (e.g., the uncreated creator). While this infinite regress may be a scientific or philosophical problem, we argue that commonsense causal reasoning is rarely faced with the consequences of this kind of regress. What is essential in the commonsense conception of causal relations is the belief that there is some process or mechanism mediating this relation, whether understood in detail or not

To summarize, we suggest that people's beliefs about causal relations include (1) a notion of force or necessity, (2) a belief in a causal process that takes place between a cause and an effect, and (3) a set of more or less elaborated beliefs about the nature of that mechanism, described in theoretical terms. Our emphasis on mechanism beliefs as a fundamental component of the conception of cause contrasts with analyses of causation framed primarily in terms of covariation. In describing the mechanism approach, it is useful to contrast these two views of causal concepts.

## **Regularity View**

The basic tenet of the regularity view is that our knowledge of a causal relation arises “when we find that any particular objects are constantly conjoined with each other”(Hume, 1777 / 1975, p. 27). For instance, a person might observe that whenever she eats shellfish, she gets sick. Then, the person might want to conclude that eating shellfish causes her to get sick. In one of the most elaborate psychological models taking the regularity view, Cheng (1997; Cheng & Novick, 1992) proposes that causal strength is a function of a covariation index,  $P = P(E|C) - P(E|\neg C)$ , where  $P(E|C)$  is the probability of obtaining the target event in the presence of a causal candidate and  $P(E|\neg C)$  is the probability of obtaining the target event in the absence of the causal candidate.

Psychological theories and models based on the principle of covariation have generally focused on how people identify causes and how people induce novel causal relations rather than what it means (or what people think it means) for a relation to be causal. Nonetheless, such theories may contain at least implicit characterizations of causal concepts. We will first examine the conception of causal mechanisms based on the notion of covariation, and discuss how it departs from the mechanism view. Then we will discuss how the two views differ in their emphasis on specific instances versus patterns or collections of cases.

Recently, the covariation approach has focused on how to represent causal mechanisms. The idea is that mechanisms can be represented in terms of a complex web of covariation, or more specifically, as a directed graph in which nodes representing variables are connected with arrows indicating causal directions (Glymour, in this volume; Glymour & Cheng, 1998; Pearl, 1996; Waldmann & Martignon, 1998). Glymour (1998), for instance, uses the following example from Baumrind (1983) to illustrate this point:

The number of never-married persons in certain British villages is highly inversely correlated with the number of field mice in the surrounding meadows. [Marriage] was considered an established cause of field mice by the village elders until the mechanisms of transmission were finally surmised: Never-married persons bring with them a disproportionate number of cats.

Glymour proposes that B is a mechanism for a correlation between A and C, if, conditional on B, the correlation of A and C goes to zero. In the above example, one observes that as the number of unmarried persons increases, the number of mice in town decreases. Conditional on the number of cats, however, the covariation between the number of unmarried persons and the number of mice would be greatly reduced. Therefore, cats are a mechanism underlying the covariation between marriage and the number of mice. Glymour represents the causal mechanism underlying this contingency as follows:

$$\# \text{ unmarried persons} \rightarrow \# \text{ cats} \rightarrow \# \text{ mice}$$

That is, single people bring in a lot of cats, which leads to reduction in the number of mice.

To give another example by Glymour (1998), consider the correlation between yellow fingers and the later occurrences of lung cancer. Having yellow fingers is not a direct cause of occurrences of lung cancer because this correlation disappears conditionalized on smoking. The mechanism behind the covariation is a common cause; Smoking caused yellow fingers and smoking caused lung cancer. This way, the claim is that causal mechanisms can be represented in terms of conditional probabilities. For this reason, Glymour (1998; also Glymour & Cheng, 1998) argues that separating mechanisms and covariations is implausible.

In one respect, we would agree with this claim; people's ideas about mechanisms may support or produce expectations about patterns of association. These expectations about patterns of covariation could be used to test hypotheses about potential mechanisms. So if someone believes that getting sneezed on causes illness via the mechanism of the transmission of germs, they should expect that the covariation between sneezing and illness is conditional on the transmission of germs. Observing a different pattern of associations might lead them to revise their hypothesis about the mechanism of illness causation<sup>2</sup>. Thus, we argue that association is related to mechanism

---

<sup>2</sup> Of course whether they see the evidence as forcing a revision and how they might revise their conception (e.g., abandoning the idea of germs, postulating a third variable) is undetermined by the data. See Koslowski (1996) following Quine(1969).

as evidence is to theory. That is, mechanism explains association as theory explains evidence<sup>3</sup>.

The point of disagreement, however, is that Glymour argues that patterns of covariation are mechanisms, and not just evidence for them. Because of that, "the separation of mechanisms and associations is very odd and implausible, and, to the contrary, it seems that an important part of learning causes might very well be learning mechanisms from associations (Glymour, 1998, p. 43)". However, we argue that the core of the concept of mechanism is a belief about a process of transmission of causal influence. We do not believe such conceptions are fully captured by the formalisms contained in covariation theories of causal reasoning.

To illustrate our point, let us go back to the previous example of the relations among the number of singles, mice, and cats. The proposal is that cats are a mechanism mediating between singles and mice because conditional on the number of cats, the relation between the number of singles and the number of mice is independent. One possible interpretation of this formulation is that *any* pattern of association in which the third factor "screens out" a correlation between two factors is a mechanism. However, there are identical patterns of data which do not involve mechanisms. Consider the variable "number of rodents." Conditional on the number of rodents, there is no relation between the number of singles and the number of mice. Yet the number of rodents is not a mechanism because the elements of a causal relation must be logically distinct. Clearly it is simple enough to modify the covariation account to stipulate that causes (mechanisms) and effects must be logically independent. However, the significance of this example is the demonstration that conditionalized covariance is not sufficient for the identification (or definition) of mechanism. .

---

<sup>3</sup> Cheng (1997) proposes a measure called "causal power," which is a contingency between two observed factors in the absence of all alternative causes. She claims that the relation between this conditionalized contingency and the observed contingency (which is not conditionalized over all alternative causes) is like the one between a theory and a law. We instead view this relation as one between a law and data. A law specifies regularity among factors in the absence of all confounding variables. Hence, a law is like the conditionalized contingency, or what Cheng terms as "causal power." The observed contingency would be like observed data which always contain potential confounds.

For a second non-example, consider the volume of cat food sold in a community. There would be a strong negative correlation between the volume of cat food sold and the number of mice, and there would be a strong positive correlation between the volume of cat food sold and the number of singles in town. Furthermore, conditional on cat food sales, the number of singles is independent of the number of mice. If the idea is that *any* pattern of association in which a covariation is screened out by the third factor is a mechanism, then volume of cat food should be also considered a mechanism underlying the relation between the number of single people and the number of mice. Yet, few people would agree with it.

We would argue that the same kind of evidence leads to the conclusion that cats are a mechanism but cat food is not, because we understand something of the causal powers of cats and cat food. Likewise, we judge that "number of rodents" is not the right sort of thing to serve as a mechanism. We argue that patterns of association and covariation are interpreted in light of beliefs about mechanisms and causal powers that are fundamental elements of conceptions of causal relations. That is, not all conditionalized covariations are considered mechanisms.

Of course it is often possible to imagine further covariation information which would rule out spurious mechanisms like the ones we just discussed. Maybe, the positive correlation between the volume of cat food sold and the number of single people is novel information, and for that reason people might want to look more carefully for other covariates. For example, we could see that cat food sale is not a true mechanism between the number of mice and the number of single people because it could be screened out by another variable, namely the number of cats. The problem, however, is that determining how to explore current covariations would be impossible without any appeal to prior knowledge about mechanisms. That is, without prior background knowledge, this process will take unlimited computing time because can only be achieved through an exhaustive covariational analysis over all objects in the world. On a purely correlational approach we won't know where to look for it. If the covariational analyses are incomplete, the

---

choice between models (i.e., determining which pattern of association is a mechanism and which one is not) cannot be based, *solely*, on covariation information.

Glymour (1998) also hints some need for existing knowledge in selecting the right models among all possible ones. If we need existing knowledge to begin with, however, the covariation account faces something of an infinite regress: some biases or predispositions must be needed to get things going. Proponents of the covariation approach often complain that the mechanism approach does not provide any account of where mechanism knowledge comes from in the first place (e.g., Cheng, 1997). But as we illustrated in the above, the same shortcoming applies to the covariation approach. Furthermore, our claim is not just that existing knowledge helps. Our claim is stronger in that it is essentially the existing causal mechanism knowledge that does all the work. We suggest that much of the disagreement between mechanism and covariation accounts stems from the accounts of the background knowledge people bring to bear in reasoning about a new case. Is prior knowledge beliefs about mechanisms, or is it the results of previous covariation analyses? Later in this chapter we will discuss this issue in more detail.

A second way in which the associationist account departs from the mechanism view is that according to the former, cause is primarily defined over samples of multiple events (general cause, henceforth). In contrast to this, we argue that it is the conception of the individual or specific case that is fundamental for commonsense conceptions of cause because the sense of transmission of causal power in an individual instance is essential.

Take the case of the live polio vaccine (known as the “oral polio vaccine,” which is given as drops in the mouth). In most cases, the vaccine prevents polio. In other cases, however, the live vaccine fails and leaves a person unprotected. In still other cases, the vaccine, because it is alive, actually causes the disease. Assume these outcomes occur randomly. If a person gets polio after receiving the vaccine and being exposed to the virus, it seems to be a matter of fact whether the vaccine caused the disease, or merely failed to prevent it. That is, two potential mechanisms of a causal outcome are distinct in this case; in one mechanism, causal power is transmitted from the vaccine, and in the other mechanism, it is transmitted from the virus that the person was exposed to

after the vaccine. Yet, from the perspective of the correlation between vaccine and disease outcome, there seems to be no difference. No pattern of covariation could distinguish which of the two mechanisms was responsible in this specific case<sup>4</sup>.

Salmon (1984) gives another example. Suppose a golfer tees off. The shot is badly sliced, but by accident, it hits a tree branch, and drops into the hole for a spectacular hole-in-one. Hitting a tree branch usually prevents a ball from going in the hole. Hence, based on a covariation analysis, hitting a tree would not be identified as a cause of successful golf shots. However, in this specific case, we have no trouble accepting that this player made a hole-in-one because the ball hit the branch.

Of course, the way to address the problem of "low probability" causes is to transform the representation of the event into a "high probability" cause. In the golf example it could be argued people are really reasoning about the causal consequences of one object hitting another. From past experience we know that "A small object hitting a large one at angle  $\theta$  causes the smaller to rebound at angle  $\theta'$ ." Thus it is easy to see the tree as the cause of the ball's travelling in a particular direction, which just happens to be the direction of the hole. As in the cat food example above, the reinterpretation requires positing both that people focus on just the correct piece of covariation information while ignoring misleading data and that they have access to the right sort of prior knowledge. Again, it is not clear which one comes first, the covariational analysis or the access to the mechanism knowledge. The difference between a mechanism account and a covariation account seems to critically involve the characterization of importance of people's prior

---

<sup>4</sup> It may be that there are ways to distinguish the two scenarios (e.g, is there polio virus in the environment, is the strain causing the illness the same as the vaccine strain, etc.). However, note that it is our sense that there are different mechanisms that drives our expectations that there are some distinct patterns of covariations, not the other way around. The central point is that despite any pattern of past covariations, it is what happened in this particular instance that determines the cause. This is the same point illustrated in the golf example following.

knowledge.

Before addressing the issue of prior knowledge, it is worthwhile to mention a different sort of argument for the individual nature of causal relations. Searle (1983) argues that an ascription of intentional cause need not involve any belief that the relation holds generally. He writes:

For example, suppose I am thirsty and I take a drink of water. If someone asks me why I took a drink of water, I know the answer without any further observation: I was thirsty. Furthermore, in this sort of case it seems that I know the truth of the counterfactual without any further observations or any appeal to general laws... And when I said that my being thirsty caused me to drink the water, was it part of what I meant that there is a universal law? ... Part of my difficulty in giving affirmative answers to these questions is that I am much more confident of the truth of my original causal statement and the corresponding causal counterfactual than I am about the existence of any universal regularities that would cover the case. (Searle, 1983, p. 118)

Similarly, we argue that the core of the idea of cause in laypeople's conception is that a particular factor influences outcomes or transmits causal powers in particular cases. Whether or not a causal relation held in one occasion will do so in the future is a secondary inference that is conceptually distinct from the original belief in cause.

On the mechanism account the belief that A caused B consists of the notion that there was some influence or transfer between the two entities or events-- something particular happened. We suggest that an associationist approach characterizes the belief that A caused B to be *primarily* an expectation about a general pattern of covariation between A and B. We have tried to show, through some examples, that people may have strong intuitions about causes even in the absence of good evidence or expectations about patterns of covariation about general cases. The associationist rejoinder to these examples is that what we are calling individual or particular causal relations are really just instances of previously learned patterns of covariation. According to them, to believe that A caused B is to believe that there is a consistent relation between events of type A and events of type B. If beliefs about causal relations must be based on patterns of covariation, then clearly

the only empirical questions surrounding causal reasoning concern whether people are sensitive to particular sorts of covariation information and how they choose between, or weight, different patterns. Thus we suggest that the associationist position relies heavily on a particular account of the origins of causal beliefs, namely covariational analysis. In advancing the mechanism approach we want to take issue with the claim that all of a person's causal beliefs are based, in any psychologically significant sense, on covariation.

### **Identifying Causal Relations**

From a discussion of how people conceptualize causal relations, we now turn our attention to the other aspect of causal reasoning: how people identify or discover that two things or events are causally related. As Rey (1983) points out in his discussion of categorization, there are typically many, varied, ways of identifying an instance of a category. We believe the same holds true for identifying instances of causal relations. In this section we will consider two means by which people might identify causal relations: induction and abduction. Inductive methods are often held up as the method by which causes are identified. We will consider the basis of these claims and will argue that induction is usually inappropriate as an account of how people ordinarily identify causes. Instead we will offer an account of causal reasoning based on abduction or “inference to the best explanation.” After comparing induction and abduction, we will discuss the role of mechanism and covariation information in induction and abduction, and will end by revisiting the debate between the mechanism and covariation views.

#### **How Inductive Reasoning Works**

In the broad sense, induction is any inference made under uncertainty, or any inference where the claim made by the conclusion goes beyond the claim made by the premises (Holland, Holyoak, Nisbett, & Thagard, 1987). In this sense, induction subsumes abductive and analogical reasoning, or any non-deductive reasoning. In the narrow sense, induction is limited to inference to a generalization from its instances (Josephson & Josephson, 1994; Peirce, 1955). For instance, given that all A's observed so far are B's, one might expect that in general all A's are B's.

We will use the narrow sense of induction, as in existing models of causal induction (e.g.,

Cheng, 1997). In a typical causal induction model, the input to the system is a set of specific instances and its output is an abstract causal rule. In a nutshell, the basic inductive learning mechanism in these models is to tabulate frequencies of specific cases in which a target effect does or does not occur in the presence or absence of a candidate cause, and to calculate causal strengths based on the contingency. This way, causal induction models can account for how causal beliefs originate from non-causal data.

### **How Abductive Reasoning Works**

Causal induction is not the whole story in causal reasoning. People not only attempt to induce novel causal relations from specific cases, they also attempt to explain why something happened in a specific case by applying known causal rules. Peirce (1955), who first described abductive inference, explains this process as follows: given an observation  $d$  and the knowledge that  $h$  causes  $d$ , it is an abduction to conclude that  $h$  occurred. For example, a detective trying to determine the cause of someone's death would proceed by generating hypotheses and assessing their fit with the existing evidence. For example, the detective might first consider whether the victim was shot (knowing that gunshots often cause death). The absence of an external wound would tend to impugn this hypothesis. Another hypothesis might be that the victim was poisoned. The detective would gather data both consistent (the victim was eating before death) and inconsistent (other people who ate the same food did not die) with this hypothesis. Eventually, perhaps after several rounds of hypothesis generation and evidence collection, the detective might decide that only one possible cause of death matches the facts of the case. Thus abduction is often called the inference to the best explanation, because in general, there are several hypotheses which might account for the evidence, and out of these, the best one is selected. Note that in abductive reasoning, the pattern of data alone does not warrant the inference. The abductive conclusion is only reached by using the data to decide among a set of alternatives generated based on existing knowledge.

The conclusion drawn as the result of an abduction ( $h$  in the above illustration) contains a vocabulary not used to describe the data ( $d$  in the above illustration). For instance, a patient with

jaundice might be diagnosed with hepatitis, which is from the vocabulary of diseases and not the vocabulary of symptoms. Because of this, abductions are often considered as "leap from observation language to theory language" (Josephson & Josephson, 1994, p. 13). Naturally, the mechanism view is compatible with abductive reasoning. Our existing knowledge about causal mechanisms is used to determine the best explanation for a given situation, just like a theory explains evidence.

Abductive reasoning is not deductively valid. It is actually an example of the classical fallacy of "affirming the consequent." Although normative criteria have been proposed in causal induction (e.g., Glymour & Cheng, 1998), no such framework has been provided for abduction. Still, uses of abductive reasoning are manifold (e.g., Harman, 1965). Examples can be found in medical diagnosis (e.g., Peng & Reggia, 1990), legal decision making (e.g., Pennington & Hastie, 1986), and even vision (e.g., Charniak & McDermott, 1985). Unfortunately, the issue of abductive reasoning has rarely been studied in Cognitive Psychology.

Generally, it is thought that judgment of the best explanation will be based on considerations such as which explanation is simpler, which is more plausible, which explains more, which fits more coherently with existing knowledge, and so on (also see Brewer, Chinn, & Samapungavan in this volume). However, the criteria for judging a good explanation need not be explicitly represented nor theoretically motivated: Peirce suggests that abduction may at times be guided simply by "aptitudes for guessing right." Unlike induction, the quality of competing explanations also directly influences abductive reasoning (Josephson & Josephson, 1994). In causal induction, the role of competing hypotheses is indirect. For instance, in the power PC theory (Cheng, 1997), alternative causal candidates influence induction only by changing the conditional probability of the target effect in the absence of the causal candidate, and one does not even need to know what these alternatives are. However, abduction is more like decision making in that a reasoner makes explicit comparisons among alternative explanations in order to select the best one.

The process of abduction is one of trying out alternative explanations and seeing which one

best fits the data. There are clearly two parts to this process. One is elaborating and refining existing explanations to improve the fit to existing data. The other is collecting more data in the hope of achieving a better fit to existing explanations. Therefore, other decisions to be made in abduction include how much effort one should put into data collection versus explanation adaptation. For example, if there is a big cost to guessing wrong, the reasoner would probably want to focus on data collection. On the other hand, if the pressure is to try something quick, the reasoner would probably be inclined to squeeze the best explanation to the data and act on it.

To summarize, the reasoning processes underlying abduction are different from those underlying induction. While causal induction is for learning general or universal causal laws from non-causal data, abduction involves collecting evidence and making decisions about a specific case. We now consider the conditions under which each kind of reasoning would be invoked.

### **Induction or Abduction?**

Induction has the virtue of describing how causal knowledge may be inferred from non-causal data. That is, given experience which does not contain information about causal relations, how do we arrive at causal beliefs? Clearly, if one's goal is to give a reductive account of how causal beliefs may arise from non-causal data, induction is the preferred account. For this reason, induction is often taken as a normative approach to causal reasoning. Indeed, there are now accounts of powerful procedures for deriving causal inferences from correlational data (e.g., Spirtes, Glymour, & Scheines, 1993). The relevance of these procedures to the question of human causal inference is where there may be some debate. One focus of research is whether people are able to apply these inductive methods. Research from a number of psychological studies suggests that people can carry out inductions that satisfactorily conform to normative principles (e.g., Cheng, 1997; Cheng & Novick, 1992; Spellman, 1996). However, a different perspective on causal reasoning asks whether this is actually how people *typically* reason. While we might be able to construct scenarios in which people perform (as if they were making) inductions, how ecologically valid are these scenarios?

In our earlier work (Ahn et al., 1995), we gave undergraduate subjects descriptions of

events. The type of events varied widely from normal everyday activities (e.g., The customer paid for the bill after eating at a restaurant) to nonsense sentences (e.g., The tove gyred the mimble). The subjects' task was to ask the experimenter questions in order to explain the events. Even for nonsense sentences, participants rarely asked for information necessary for causal induction. That is, they did not engage in the process of selecting a causal candidate and collecting the information needed for a covariational analysis. Instead, most responses seemed to reflect an attempt to figure out which known causal mechanism best fit the given situation. For instance, participants frequently introduced new theoretical constructs not mentioned in the descriptions of the events. They also asked whether preconditions for a hypothesized mechanism is satisfied even for nonsense sentences (e.g., "Was the tove mad at the mimble?"). That is, they seemed to be making causal abductions-- inferences to best explanation, or inferences to mechanisms (see also Lalljee, Lamb, Furnham, & Jaspars, 1984; Major, 1980; White, 1989 for similar results).

Sometimes, one has no choice but to apply existing causal knowledge rather than collecting data necessary for causal induction. Most notably, this often happens when we need to discover what caused an individual event. By definition, covariation information requires multiple instances. At the very least, one needs to observe the consequences of presence and absence of a causal candidate. However, with only a single specific case, even this minimum requirement cannot be met. For instance, a covariational analysis can establish that cigarette smoking causes lung cancer in general. However, it cannot establish whether cigarette smoking caused a *particular* lung cancer, because one cannot observe what would have happened if the person did not smoke cigarettes.

Even with general cases, there are other reasons to believe that abduction is preferred over induction. The normative way of inducing causal relations from covariation assumes that people have a sufficiently large number of representative samples. But in real life, covariation information is not presented all at once in a prepackaged form. Rather, each observation is made one at a time in a sequential manner. Clearly, it is difficult for a reasoner to determine a priori how many observed instances make up a representative sample. Hence, one often must make causal

attributions on-line as the observations are made. Indeed, Dennis and Ahn (1997) demonstrated that people are willing to make causal attributions based on a small set of cases before waiting to see the whole array of examples. As a result, they exhibit a strong order effect depending on the type of instances they encounter first (see also Lopéz, Shanks, Almaraz, & Fernández, 1998 for another example of the order effect in causal induction). Using data as it comes in and not assuming unrealistic representational capacities can be also thought of as characteristic of evolutionarily plausible reasoning strategies (Gigerenzer & Goldstein, 1997).

The second reason why causal induction might be less prevalent than abduction in real life is that some evidence necessary for calculating conditionalized covariations is impossible to obtain. Recall the previous example of correlation between increase in the amount of cat food sold and decrease in the number of mice. In order to determine whether cat food sales directly cause the number of mice, one might want to calculate the contingency between these two while holding the number of cats constant. Practically speaking, however, such evidence would be impossible to obtain. Indeed, there are numerous other factors (e.g., cat stomachs, cat tails, etc.) whose covariations are (practically) impossible to observe while holding the number of cats constant. In fact, in real life situations, reasoners normally do not have the luxury of obtaining all the evidence they would like to see before making inductive conclusions.

Although nobody has explicitly claimed that causal induction is the only way people identify causes, the current focus of causal reasoning models gives the impression that causal induction is more important and fundamental than abduction. Contrary to this, we claim that abduction may be the most pervasive and natural means of identifying causes. One might discount this point by saying that it is obvious that once causal knowledge is induced, one might as well use it rather than trying to learn new knowledge afresh (Glymour & Cheng, 1998). However, our claim goes beyond this. In explaining why abduction might be more prevalent than induction, we alluded to the problems that are inherent in causal induction. That is, abduction is preferred to induction not simply because of cognitive efficiency, but because most of the time, a complete covariation analysis (not guided by prior mechanism beliefs) is impossible.

## **Learning**

Abduction does not answer the question of the origins of causal beliefs. However, it should be noted that causal induction is only a single (though important) aspect of learning new causal relations. That is, we should not identify causal induction with the question of how people acquire new causal beliefs or how people identify causes. Some causal knowledge is innate; nobody teaches a baby the consequence of sucking behavior. Some could be acquired through analogical reasoning. For instance, Read (1983) has demonstrated that people make causal predictions even from a single example especially when the causal rule is complex. Of course, the most common and efficient way of acquiring causal knowledge would be through direct instruction or communication (see Sperber, 1996 for more detailed discussion of these processes). There are also more dubious ways to arrive at the conclusion that one thing caused another, such as dreams or visions. As with any other type of belief, people may come to causal beliefs in a host of ways.

In some respects, one might argue that non-inductive accounts for learning are circular: In saying that causal knowledge is the result of communication, how did the person who transmits the belief come to have it in the first place? If one traces back to the origin of basic causal beliefs, didn't they come from a covariational analysis? As we already discussed, it is not difficult to think of ways other than covariational analyses. Some causal beliefs could be innate. Some could have been fabricated. On the other hand, the covariation-based induction method might actually be most responsible for the creation of new causal beliefs. Even in that case, however, it is not clear what bearing it has on the transmission of causal beliefs. For example, person A might acquire from person B the belief that unprotected sex is a cause of AIDS. The basis for person B's belief, be it an analysis of covariation or a particular religious conviction, may be completely unknown to A. In this case it would seem odd to suggest that the bases of B's beliefs have any significance for A. That is, induction might account for the origin of causal beliefs, but that does not undermine the argument that non-inductive learning methods are prevalent in most situations of learning new causal relations.

## **Summary**

To summarize, this section described two ways of identifying causes: induction and abduction. We first explained that these two methods are distinct reasoning processes. Then we argued that in everyday causal reasoning, abduction is more prevalent than induction. We described empirical results for this claim along with a theoretical analysis of why this would be the case. One counter-argument for our claim is that abductive reasoning cannot explain the learning of novel causal relations. Although we acknowledged that there is some truth to this, we argued that learning new causal relations can be carried out through means other than causal induction.

## **Role of Covariation and Mechanism Information in Induction and Abduction**

Traditionally, the covariational models have focused on causal induction while the mechanism view has focused on abduction. However, we do not argue that covariational information is useless in abductive reasoning. At the same time, although the mechanism view has often been criticized for lacking an account of the origins of mechanism knowledge (Cheng, 1997), we believe that mechanism information is indispensable in induction. In this section, we discuss how mechanism information contributes to induction and how covariation information contributes to abduction.

### **Use of Covariation Information in Abduction**

Covariation information is certainly useful in abduction, although not necessary. Suppose John Doe has lung cancer and there are three possible explanations for this case; he smoked, his family has a history of lung cancer, and he worked in a stressful environment. In determining which one is the best explanation, the base rate (or some kind of covariation-based index) of these candidate factors in the general population would certainly be useful. It should be noted, however, that covariation is only one type of data in abduction. Another important type of data would be the fit between hypothesized mechanisms and specific cases. In the above example, one might also want to know how much he smoked, how close to him was his relative with lung cancer, and so on. Sometimes, these other kinds of evidence can outweigh the evidence provided by covariation,

as in the case of base rate neglect due to representativeness heuristics (Kahneman & Tversky, 1973; see also our previous discussion on the specific versus general causes). Therefore, although covariation information is beneficial, it is neither sufficient nor essential for abduction.

### **Use of Mechanism Information in Induction**

There is now a general consensus that inductive reasoning requires constraints because of computational complexity (e.g., Keil, 1981; see also the special issue of *Cognitive Science*, 1990 which is devoted to this issue). Clearly, the principle of association is one of the most fundamental learning mechanisms. But the need for additional constraints has been thoroughly recognized in various domains, including language acquisition (e.g., Chomsky, 1965), concept learning (e.g., Murphy & Medin, 1985), and even learning of covariation per se (e.g., Alloy & Tabachnik, 1984). We re-introduce this rather widely accepted idea here because it has not been specifically discussed in the context of causal reasoning. Furthermore, the recently developed covariation-based approach to causal mechanisms seems to be neglecting this issue. Our main claim is that while covariation-based learning of new causal rules is essential, induction is almost impossible without our existing knowledge of causal mechanisms, because there are simply too many possibilities<sup>5</sup>. Below we describe some of the specific ways mechanism information can guide learning new causal rules.

Determinants of causal candidates. In order to determine causal candidates for covariational analyses, one must start out with some understanding of causal mechanisms because otherwise, one will soon run into computational explosion (see also Peirce, 1955). This is often called the frame problem in artificial intelligence. The reason why a normal person would not even think about the possibility that wearing red socks might cause Alzheimer's disease in their later life is precisely because we cannot think of any plausible mechanisms. To quote Popper (1963),

---

<sup>5</sup> This statement should not be taken as the claim that “people don’t learn causes from associations” or that “causes have nothing to do with associations” as Glymour (1998, p. 41) misconstrues it. Our claim (in this chapter, as well as in our previous work) is that associations alone are insufficient for learning causal knowledge, and that there are many ways of learning causal knowledge other than associations.

The belief that we can start with pure observations alone, without anything in the nature of a theory is absurd ... Twenty five years ago I tried to bring home the same point to a group of physics students in Vienna by beginning a lecture with the following instructions: Take a pencil and paper; carefully observe, and write down what you have observed. They asked of course, what I wanted them to observe ... Observation is always selection. It needs a chosen object, a definite task, an interest, a point of view, a problem. (p. 46)

We do not deny the possibility of starting out with observations of unexplained correlations and then imposing causal interpretations on them. However, this type of purely bottom-up covariational analysis seems extremely rare in real-life situations. As described in the previous section, we have observed it to be a rare exception in our earlier work (Ahn et al., 1995). A true discovery of new mechanisms from observations can only be achieved by a few scientists whose job is devoted to this kind of task. In most everyday reasoning, people start out with hypotheses on causal candidates that are generated from their existing mechanism knowledge.

Relevancy of data. Even after we select a manageable set of causal candidates to test for a covariational analysis, a reasoner still has to decide which events are relevant for the analysis. In a typical covariational analysis, four pieces of information are needed, crossing presence / absence of the causal candidate with presence / absence of the target effect. The difficulty of a covariational analysis arises in particular when one needs to decide what counts as the absence of an event. The worst possible case would be the joint absence of the causal candidate and the target effect (Einhorn & Hogarth, 1986). As discussed in the famous “ravens paradox,” although "All ravens are black" is logically equivalent to "All non-black things are non-ravens," our observation of a purple flower does not strengthen our belief that all ravens are black. For any covariational analysis there is an infinite possible number of joint absent cases, but they cannot possibly all strengthen our causal beliefs. Only some of them seem relevant in increasing our causal beliefs.

Salmon (1966) discusses the problems that accompany this arbitrariness of probabilities. The idea is that the larger the reference class, the more reliable the statistics, but the less relevant.

---

For instance, in estimating the likelihood that Michael Jordan will catch a cold this winter, we might estimate the frequency of colds based on all the people in the world over the entire history of humankind. Although these statistics might be reliable, it might be more relevant if the estimate is based on male basketball players just during winter and maybe just for this year, since the dominant type of cold virus changes yearly. Note that while determining the relevancy, causal interpretation is already starting to creep in (Josephson & Josephson, 1994). That is, without constraints from existing causal mechanism knowledge, probability information can be vacuous and inapplicable.

Interpretation of data. Wisniewski and Medin (1994) convincingly demonstrated that people's domain theories determine how features are interpreted in categorizing objects. In this study, participants received children's drawings that were described as belonging to two different categories. Participants' interpretation of features was heavily dependent on the category labels provided to the drawings. For instance, a circular configuration of lines in a drawing was described as a "purse" when participants were told that a "city child" had drawn it. But when other participants were told that the same drawing was done by a "creative child," the same feature was described as a "pocket" and was interpreted as evidence that the drawer paid attention to detail. Similarly, a single observation can lead to different causal conclusions depending on the underlying beliefs (including beliefs about mechanism), which determines how the given observation is perceived and interpreted. For example, suppose one observes that the common cold is cured everytime she had chicken soup. Depending on prior beliefs, she can interpret these observations as support for a hypothesis that eating something warm cures the common cold, or as support for an alternative hypothesis that special ingredients in chicken soup cures the common cold.

When do we doubt correlations? Covariations cannot be equated with causal relations. To deal with this problem, a recent approach is to calculate causal strengths based on conditionalized covariations as implemented in the Power PC theory (Cheng, 1997). For instance, one might notice that birth defects tend to occur among mothers who use computers. But if the contingency is

re-calculated holding alternative factors constant, the previously positive contingency might disappear. For instance, if one calculates covariation between birth defects and use of computers in the absence of job-related stress, the covariation might become near zero. In that case, the contingency between birth defects and the use of computers is considered a spurious correlation and not a causal relation. However, covariation of a real cause will not be screened out conditional on alternative causes. For instance, if the covariation between birth defects and job-related stress remains positive in the absence of high alcohol consumption (or any other alternative causes), we can conclude that job-related stress is indeed a cause of birth defects. Therefore, according to the Power PC theory, "covariation does imply causation when alternative causes are believed to occur independently of the candidate (p. 374)."

An important question that has not been answered is: how do people know a priori which correlations might be spurious, that is, under what conditions would they be forced to calculate further conditional covariations? Sometimes an observed correlation is readily accepted as a causal relation without having to calculate conditional covariations, but sometimes it is not. For instance, there has been a report of a positive correlation between who wins the Superbowl (AFC or NFC) and whether the stock market goes up or down, but no one would draw a causal link between these two. On the other hand, a report on a correlation between the use of aluminum foil and Alzheimer's disease created a hype several years ago. The critical difference between these two examples seems to be whether or not people can conjecture a plausible mechanism underlying the correlation.

Sometimes, one might even dismiss correlation data per se because the reasoner cannot think of any plausible mechanisms. For instance, consider the following statistics from astrology (Gauquelin, 1967); There are more schizophrenics born in winter months than in summer. Children born in May, June, September, and October have higher IQ scores than those born in other months. Those who were born immediately after the rise and culmination of Mars and Saturn are more likely to be physicians. Most people educated under modern science would dismiss such data, let alone draw causal links. Now, consider that it has further been discovered that it's not just the children's but also the parents' birth dates that are correlated with the children's careers.

(Indeed, these are all "real" correlations reported in Gauquelin, 1967.) No matter how many more complex covariation patterns we discover, however, most well-educated people would dismiss these findings because they do not believe there could be a mechanism linking planets to personality.

However, Gauquelin (1967) presents an intriguing mechanism-based explanation for these correlations. There are three elements to this mechanism. First comes the proposition that personality (and hence career) and intelligence are genetically determined. Second is the suggestion that the human fetus is sensitive to subtle gravitational forces (for example, marine animals show exquisite sensitivity to minute tidal forces caused by celestial bodies). Finally, add the fact that the fetus precipitates labor, and you have the beginnings of an account. Fetuses with particular personality attributes signal labor in response to particular tidal forces. Once some causal mechanism has been provided, the correlations begin to warrant serious consideration in a way that no additional amount of data would accomplish.<sup>6</sup> Power PC theory, however, is silent about how *credibility* of covariation information might be determined as a function of whether or not covariation information is consistent with existing mechanism knowledge

Induction of causal mechanism. So far, the discussion has been limited to the role of mechanism information in induction of single-layer or one-step causal links. Little is known about whether people can actually learn multi-layered mechanisms only from covariation. The only empirical study known to us gives a pessimistic picture. Hashem and Cooper (1996) generated nine sets of relatively simple causal networks (e.g.,  $A \rightarrow B \rightarrow C$ , or  $A \leftarrow B \rightarrow C$ ) instantiated as diseases. Twenty second and third year medical students were instructed to ask for any conditional probabilities among the three variables in each network, and to estimate the causal strength between B and C after receiving answers to their questions. Even from these simple causal networks, their

---

<sup>6</sup> Note the proposal of a mechanism suggests the kinds of relevant covariation data. In particular, given this account we'd like to see whether the effects are conditional on natural or caesarian birth.

estimates significantly deviated from the normative answers. Although exploratory, the results suggest the need for additional constraints for learning causal networks from the bottom-up. Indeed, Waldmann and Martignon (1998), who make use of a Bayesian network to represent mechanism knowledge, admit that it is improbable that humans learn such networks bottom-up, as instantiated in some computational models (e.g., Spirtes et al., 1993).

### **The Covariation versus Mechanism Debate Revisited**

Various issues have been discussed throughout this chapter. In the final section, we compare the covariation-based and the mechanism-based approaches once again and summarize their points of agreement and disagreement.

Both views agree that our prior background knowledge about causal relations plays a role in causal reasoning. Neither approach denies abductive reasoning. The major discrepancies between the two views lie in their differing conceptions of how people think about causal relations and causal mechanisms, and in their differing emphasis on the role of mechanism knowledge in identifying causes. The regularity view does not consider people's beliefs in a necessary force or causal power, and in its current form, it proposes that conditionalized covariations imply causality. We argue (1) that conditional covariations are not what people think of as causal relations and, (2) that calculating conditionalized covariations is not how people typically identify causal relations. The pattern of association alone cannot determine which model is a causal mechanism. In real life situations, induction is impossible without the guide of existing mechanism knowledge. Hence, even if one could develop the most accurate and normative model of causal induction, which could learn complex causal mechanisms from scratch, there still would be a tremendous gap between such a model and a psychological model of everyday causal reasoning.

Our discussion of the significance of mechanism information in induction should not be taken as a direct refutation of covariation-based models of causal induction. Some of these problems (e.g., how to determine causal candidates) have been explicitly acknowledged as issues outside the realm of causal induction models. We do not attempt to dismiss the value of causal induction models. Indeed, the development of models to represent the conditions under which

covariations can be equated with causality constitutes an impressive accomplishment of the regularity view.

At the same time, it is also important to realize limits of the inductive models. Glymour (1998) argues that separating mechanism knowledge from covariations “puts everything on a false footing.” We believe that it is rather the normative approach to everyday causal reasoning that puts everything on a false footing. As our discussion has shown, human cognitive capacity and the surrounding environment simply do not provide the circumstances necessary for such a normative model to operate in the real world.

## **Acknowledgement**

We would like to thank Lawrence Barsalou, Martin Dennis, and Douglas Medin for their helpful comments on the earlier version of this chapter. We also would like to thank Jennifer Amsterlaw for her thorough proofreading of the manuscript. This work was supported by a National Science Foundation Grant (NSF-SBR 9515085) and by a National Institute of Mental Health (RO1 MH57737) given to Woo-kyoung Ahn, and by a grant from the Agency for Health Care Policy and Research (RO3 HS09556) to Charles Kalish.

Correspondence concerning this article should be sent to Woo-kyoung Ahn, Department of Psychology, Yale University, P.O. Box. 208205, New Haven, CT 06520, [ahn@pantheon.yale.edu](mailto:ahn@pantheon.yale.edu) or to Chuck Kalish, Department of Educational Psychology, University of Wisconsin-Madison, 1025 W. Johnson St. Madison, WI 53706, [cwkalish@facstaff.wisc.edu](mailto:cwkalish@facstaff.wisc.edu).

## References

- 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.
- Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior experience and current situational information. Psychological Review, 91, 112-149.
- Baumrind, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. Journal of Personality and Social Psychology, 45, 1289-1298.
- Charniak, E., & McDermott, D. (1985). Introduction to artificial intelligence, Reading, MA: Addison-Wesley.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. Psychological Review, 104, 367-405.
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. Cognition, 40, 83-120.
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. Psychological Review, 99, 365-382.
- Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge: MIT Press.
- Dennis, M. & Ahn, W. (1997). Order of confirmatory evidence affects judgment of causal strength. Poster presented at the 38th Annual Meeting of The Psychonomic Society, Philadelphia, PA.
- Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. Psychological Bulletin, 99, 391-397.
- Gauquelin, M. (1967). The cosmic clocks: from astrology to a modern science : Henry Regnery Company.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review, 103, 650-669.

Glymour, & Cheng, P. W. (1998). Causal mechanism and probability: A normative approach. In M. Oaksford & N. Chater (Eds.) Rational models of cognition. Oxford University Press.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In S. A. G. Lawrence A. Hirschfeld (Ed.), Mapping the mind: Domain specificity in cognition and culture. (pp. 257-293): Cambridge University Press, New York, NY, US.

Harman, G. H. (1965). The inference to the best explanation. Philosophical Review, 74, 88-95.

Harré, R. (1988). Modes of explanation. In Hilton, D. J. (Ed.), Contemporary science and natural explanation: commonsense conceptions of causality, Harvester Press, Brighton, Sussex.

Harré, R., & Madden, E. H., (1975). Causal powers: A theory of natural necessity. Totowa, NJ: Rowman and Littlefield.

Hashem, A. I., & Cooper, G. F. (1996). Human causal discovery from observational data. Proceedings of the 1996 Symposium of the American Medical Informatics Association .

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. Psychological Bulletin, 118 (2), 248-271.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. Psychological Review, 93 (1), 75-88.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1987). Induction . Cambridge, MA: MIT Press.

Hume, D. (1777/1975) An enquiry concerning human understanding. Oxford University Press.

Josephson, J. R., & Josephson, S. G. (1994). Abductive inference . Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80 (4), 237-251.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. Psychological

Review, 88 , 197-227.

Koslowski, B. (1996). Theory and evidence. Cambridge, MA: The MIT Press.

Lalljee, M., Lamb, R., Furnham, A. F., & Jaspars, J. (1984). Explanations and information search: Inductive and hypothesis-testing approaches to arriving at an explanation. British Journal of Social Psychology, 23 (3), 201-212.

Lopéz, F. J., Shanks, D. R., Almaraz, J., & Fernández, P. (1998). Effects of trial order on contingency judgments: A comparison of associative and probabilistic contrast accounts. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 672-694.

Mackie, J. L. (1974). The cement of the universe: A study of causation . London: Oxford University Press.

Major, B. (1980). Information acquisition and attribution processes. Journal of Personality and Social Psychology, 39 , 1010-1023.

McGill, A. L. (1989). Context effects in judgments of causation. Journal of Personality & Social Psychology, 57 (2), 189-200.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. Psychological Review, 92 , 289-316.

Pearl, J. (1996). Structural and probabilistic causality. In D. R. Shanks, D. L. Medin, & K. J. Holyoak (Eds.), Psychology of learning and motivation, Vol. 34: Causal learning (pp. 393-435). San Diego, CA: Academic Press.

Peirce, C. S. (1955). Abduction and induction. In J. Buchler (Ed.), Philosophical writings of Peirce (pp. 150-156). New York: Dover.

Peng, Y., & Reggia, J. A. (1990). Abductive inference models for diagnostic problem solving . New York: Springer-Verlag.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. Journal of Personality & Social Psychology, 51 (2), 242-258.

Popper, K. R. (1963). Conjectures and refutations: The growth of scientific knowledge . New York: Harper & Row.

Quine, W. V. (1969). Natural kinds. in W. V. Quine, Ontological relativity and other essays. New York: Columbia University Press.

Read, S. J. (1983). Once is enough: Causal reasoning from a single instance. Journal of Personality & Social Psychology, 45 (2), 323-334.

Rey, G. (1983). Concepts and stereotypes. Cognition, 15, 237-262.

Salmon, W. C. (1966). The foundations of scientific inference . Pittsburgh: University of Pittsburgh Press.

Salmon, W. C. (1984). Scientific explanation and the causal structure of the world . Princeton, NJ: Princeton University Press.

Searle, J. R. (1983). Intentionality. NY: Cambridge University Press.

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. Psychological Science, 7 , 337-342.

Sperber, D. (1996). Explaining culture: A naturalistic approach . Oxford: Blackwell Publishers.

Spirtes, P., Glymour, C., & Scheines, R. (1993). Causation, prediction, and search . New York: Springer-Verlag.

Waldmann, M. R., & Martignon, L. (1998). A Bayesian Network Model of Causal Learning. Proceeding of the 20<sup>th</sup> Cognitive Science Conference, Hillsdale, NJ: Erlbaum.

White, P. A. (1989). A theory of causal processing. British Journal of Psychology, 80 , 431-454.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning, Cognitive Science, 18, 221-281.