Learning Causes: Psychological Explanations of Causal Explanation[1]

CLARK GLYMOUR
University of California at San Diego and Carnegie Mellon University

Abstract. I argue that psychologists interested in human causal judgment should understand and adopt a representation of causal mechanisms by directed graphs that encode conditional independence (screening off) relations. I illustrate the benefits of that representation, now widely used in computer science and increasingly in statistics, by (i) showing that a dispute in psychology between 'mechanist' and 'associationist' psychological theories of causation rests on a false and confused dichotomy; (ii) showing that a recent, much-cited experiment, purporting to show that human subjects, incorrectly let large causes 'overshadow' small causes, misrepresents the most likely, and warranted, causal explanation available to the subjects, in the light of which their responses were normative; (iii) showing how a recent psychological theory (due to P. Cheng) of human judgment of causal power can be considerably generalized: and (iv) suggesting a range of possible experiments comparing human and computer abilities to extract causal information from associations.

Key words: cause, causation, directed graphs, explanation, judgment, under certainty.

1. The Puzzles

The abstract of David Shanks' recent Experimental Psychology Society Prize Lecture (1995) contains a hidden puzzle:

> We can predict and control events in the world via associative learning. Such learning is rational if we come to believe that an associative relationship exists between a pair of events only when it truly does.

We get around in the world because we know what events will follow actions, whether our own or others; we understand and can explain the world because we know something about the processes that produce what we observe. The capacity to learn what causes what and when, with all that implies about categorization, is surely among the most fundamental cognitive abilities we have. How do we do it? Shanks' abstract, and the essay that follows it, suggests a part of an answer: we learn about causes by learning about associations among types of events or types of processes. Leaving aside the particulars of Shanks' theory, his answer – we learn about causes by observing associations – is, I believe, correct. The puzzle is how it could possibly *be* the correct answer.

I am sure that, like everyone else, Shanks learned that correlation is not causation, but his second sentence collapses the distinction and confounds learning associations with learning how to predict and control. Knowing only the association

Minds and Machines **8:** 39–60, 1998.

between *A* and *B* doesn't usually enable us to control either *A* or *B*. The association between yellowed fingers in youth and middle age and lung cancer in later life doesn't of itself provide a way to control lung cancer. Preventing yellow fingers will do nothing to change the frequency of lung cancer unless the intervention also changes smoking frequency – making everyone wear gloves, for example, will do nothing to reduce lung cancer rates. 'Prediction' of *B* from *A* is ambiguous: we may predict B *given the occurrence* of *A* or we may predict *B given* an *intervention* to bring about *A*. The two are not the same, and need not issue in the same probabilities, as a little reflection on the yellow fingers/lung cancer example will show. So there is the puzzle, or rather the set of puzzles: how could *any* cognitive system learn causal relations from associations? How do people do it? Are there experiments that might answer the second question?

There is, of course, absolutely nothing new in these questions. The first two have been the cynosure of philosophical accounts of inquiry since Aristotle, and great modern philosophers – Descartes, Leibniz, Hume, Kant – each marked their originality with an answer. The questions were the forgotten purpose of some great works we still remember. George Boole's *The Laws of Thought*, which introduced the algebra that now describes most of digital technology, aimed above all to provide a method of inferring causes from observed effects.

Boole's own work provides a caution about these questions. So far as we know, the theory of deductive inference began with Aristotle and, through two millennia of sporadic attention, was little improved until Boole's work in the 1850s and Frege's in 1879. The more or less correct normative theory that emerged in the 19th century is complex and in many respects counterintuitive. (Lance Rips, in a lecture at Carnegie Mellon University, reported on an experiment in which undergraduate psychology students who had no training in formal logic and a comparison group of students who had previously taken a logic course were given the same problems in sentential logic. After attempting the problems, all students were given the correct answers. On a subsequent test with similar problems, the students who had formerly taken logic improved their performances, but the untrained students performed *worse* than on the first test. Moral: the correct theory is tough to learn.).

A similar lesson can be extracted from the history of probability. Aristotle wrote about chance, and phenomena of frequencies have been used in science at least since Greek astronomy. Kepler complained of the absence of a theory of random errors. But no actual probability calculations appeared until the 16th century, nothing we recognize as real probability theory until the 17th, no statistical inference until the 19th, and no axiomatization of the theory until the 20th century. And, as a wealth of psychological literature has shown in the last decades, people have trouble with probability. *There is no reason to expect a correct, normative theory of inference to causes from associations to be simple, easy, or intuitive, or to expect that human inference will accord with it except in simple problems*. The universe has no key, only a combination lock.

In what follows I describe the elements of a representation of causal explanations – a representation now almost standard in computer science and increasingly common in statistics – that provides the basis for algorithmic solutions to these puzzles; that is, mathematical work using the representation shows how (and what) causal information may be reliably extracted from observed associations, and how that usually incomplete causal information can be used in prediction and planning.

This work has at east four kinds of implications for psychology: methodological, interpretive, analytic and substantive. The methodological issues have principally to do with old fashioned but still relevant problems, such as the justification of 'intervening variables,' and with entirely contemporary issues about techniques of data analysis and theory construction (by psychologists, not their subjects). The interpretive issues have to do with understanding the confusions in false dichotomies between 'associationist' and 'mechanist' accounts of causation, confusions so influential they threaten to eliminate from psychology any serious work on the subject of causation, and with the interpretation of experiments intended to assess whether, how much, and why, human judgment of causal relations is sub-normative. The analytical issues have to do with unfolding the hidden implications of contemporary psychological theories when they are translated into the new representation. The substantive issues have to do with the main puzzle implicit in Shanks' abstract: how do humans extract the available causal information from associations?

Of the four kinds of implications, I will ignore the methodological topics in this paper, but I will illustrate each of the others. The substantive issue, which in my view is the most interesting and important, I will deal with last.

2. Mystery "Mechanism": An Answer Too Many Psychologists Like

Possibly the most popular (among psychologists: see Ahn and Bailenson, 1996; Ahn et al., 1995; Baumrind, 1983; Schultz, 1982; White 1989, 1995; for philosophers of the same opinion see; Harré and Madden, 1975; Turner, 1987) answer to the questions I extracted from Shanks' abstract denies their presupposition: people *don't* learn causes from associations, because causes have nothing to do with associations, they have to do with 'mechanisms.' What is meant by 'mechanism' is rarely explained in this literature, but the examples make it relatively clear that to specify a 'mechanism' for a covariation is simply to specify either a sequence of causes that intervene between the candidate cause and effect, or causes that tend to bring about both the candidate cause and effect, where the causal connection posited in the 'mechanism' are of a kind that are already familiar and acknowledged. Baumrind (1983) gives the following illustration:

> 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.

Similar examples are offered by Ahn et al. (1995) and others. Mechanisms of this kind can be represented by a causal diagram or *directed graph* (i.e., a network of nodes representing features or variables and with arrows pointing from causes to effects), for example

# unmarried persons $\rightarrow$ # cats $\rightarrow$ # mice

There are two sorts of probabilistic consequences to this sort of mechanism: the mechanism implies relations among conditional probabilities, and the mechanism implies probability relations upon various interventions. The two sorts of probability implications will sometimes, as in this case, be equal, but they are not the same. In Baumrind's example, if hers is the entire mechanism behind the association, then if we were to *intervene* to hold the number of cats constant in these villages, there would be no frequency association between variations in the number of unmarried persons and variations in the number of mice. And under the same assumption, if we did not intervene at all, but simply computed the conditional probability of any number of mice *given* any number of unmarried persons and any number of cats, the result would approximately equal the conditional probability of that number of mice given that number of cats. And, by almost any measure of covariation, the (negative) covariation between married persons and mice should be weaker than the (negative) covariation between cats and mice.

So suppose we knew nothing about the English habit of pet keeping, and we were ignorant of the disposition of cats towards mice, but we discovered the following *associations*:

# of unmarried persons is negatively associated with # of mice;

# of unmarried persons is positively associated with number of cats;

# of unmarried persons is negatively associated with # of mice;

and the single conditional independence:

# of unmarried persons is independent of # of mice given # of cats.

Assuming we have reason to believe that the observed associations are not produced by whatever method of measurement was used, these relationships invite only a few alternative causal explanations, which we can sketch diagrammatically:

1. #mice —-> # cats —-> # unmarried persons
2. # mice <— Something Else —-> # cats —-> # unmarried persons
3. # unmarried persons —–> # cats —-> # mice
4. # unmarried persons <—- Something Else —-> # cats —–> # mice
5. # unmarried persons <—- # cats —–> # mice

If, separately, we knew, for example, that the number of cats does not cause the number of unmarried persons, we could eliminate all but explanations 3 and 4. Both of these explanations suppose that the number of cats influences the number of mice and we could also conclude that *either* the number of unmarried persons influences the number of cats, *or* something else influences both.

So 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 together with prior knowledge. Later we will see that these inferences can be made rigorous.

Besides Baumrind's example, consider briefly the mechanism that generates the covariation between past occurrences of yellow fingers and the later occurrences of lung cancer among those who grew up in the days of unfiltered cigarettes. The mechanism behind the covariation is a common cause: smoking caused yellowed fingers and it also caused lung cancer:

$$\text{yellowed fingers} \leftarrow \text{smoking} \rightarrow \text{lung cancer}$$

Here again, there are two distinct kinds of probabilistic implications of the explanation. First, yellowed fingers and lung cancer are independent conditional on smoking. (More generally, when there are no other causal connections, the effects of a common cause are independent conditional on a value of the common cause (Simon, 1977)). Second, interventions that directly alter only the frequency of yellowed fingers do not change the probability of lung cancer.

As these examples illustrate, there are intricate connections between mechanisms and patterns of association, and a fruitful mechanistic approach to understanding both norms of causal inference and human judgment about causation might try to understand those patterns and investigate the ways humans use them, or can learn to use them, to infer causal mechanisms. But, a disconnection between mechanisms, on the one hand, and probabilistic patterns, on the other, puts everything on a false footing.

There is a kind of ecological fallacy in the mechanistic literature that mistakes the frequency of a phenomenon for its explanatory importance. For example, Ahn, et al. (1995) claim that information on covariation is generally not necessary for learning causal relations. With 'rare exceptions,' people do not learn causal relations from covariations but from applying prior knowledge of mechanisms. The idea is that people assess whether an association between A and B is causal by seeing whether A and B are instances of cause and effect for some mechanism already known to them, approximately what is called 'explanation based learning' in the artificial intelligence literature.

Perfectly plausible. But what about the 'rare exceptions'? Compare language learning: Over the history of any individual's speech production events concerned with learning the syntax, semantics and pragmatics of a language are 'rare exceptions,' and most of our inner searches for what to say in a particular context consist in applying what we already know. That does not make understanding the acquisition of first languages the less interesting or fundamental. Or, for that matter, compare terrestrial bodies. Mostly they don't fall or move, they just stay where they are. The intellectual ancestors of Ahn et al. would presumably have concluded that the study of motions of bodies must be marginal.

3. Another Answer: Conditioning

Classical and operant conditioning both produce an appropriate expectation in a learner, but they differ in what the learner discovers about control. In classical conditioning a subject learns an association between two kinds of events, neither of which are interventions or actions of the learner. In operant conditioning, a subject learns an association between two kinds of events, one of which is an action of the learner. In classical conditioning, unless there is other relevant knowledge, all that can be learned is an association, not a causal relation, because the causal process responsible for the association is underdetermined by the association itself. Pavlov's dogs could not know whether the bell ringing caused dinner, or dinner the bell ringing, or something else caused both. Neither could we in circumstances comparably bereft of information. In contrast, in most cases the proposition describing what is learned in operant conditioning is that an action of a certain sort *causes* an event of a certain kind, and the learner acquires the capacity to control, or at least to influence, the occurrence of the relevant events.

Perhaps, outside of deliberate scientific inference, that is all the causal learning that humans do, and all they need to do in the ordinary courses of life. But there are at least weak reasons to think otherwise. Piaget suggests that children string together information about actions and information about associations to identify 'remote' causal connections. And one can think of a variety of circumstances in which the capacity to correctly identify causes from associations would have promoted survival: A hunter studying a very small herd of antelope watches the herd. From their behavior alone the hunter wants to identify the lead animal because if the lead animal can be guided to an ambush the others will follow. Or again, a gatherer sees that a wind, varying with time in direction and intensity, blows on a field of tall grass. The grass moves in bunches. Is there a hidden enemy throng moving in the grass?

Perhaps there are more or less reliable ways to make such inferences and, for evolutionary reasons, we can do so. Or perhaps not. These are only so many 'perhaps' on either side. Only a systematic course of experiments comparing human and normative performance at identifying causal structure will illuminate the issues. But what norms? There is a normative theory, apparently entirely unknown to psychologists, but is only a single experimental study of these questions, and none at all in the psychological literature.

4. Representation

A normative account of causal inference requires a representation of causal relations general enough to include, to good enough approximation, the great majority of causal systems we think we encounter, and a characterization of the information about causal relations that can be extracted from observed associations, or from observed associations and prior knowledge of various kinds. It does not require an *analysis* of causation or a representation or an account of inference that covers every imaginable case. (Much of the poverty of contemporary philosophy results from insisting on perfect generality, thereby avoiding the effort of investigating any unobvious consequences of any assumptions, since none are perfectly general.) What follows in this section is not mysterious, and in many respects not even difficult. It requires some patience with formal definitions and distinctions, and some elementary modern mathematics. It is an abbreviated description of the representation of causal mechanisms that has become almost standard in computer science, and is implicitly used throughout much of applied statistics. The payoff is astonishing.

In discussing the mechanist view, and Baumrind's example in particular, I introduced diagrams with nodes indicating features of a system and an arrow, or directed edge, from one node to another indicating that the feature represented by the node at the tail of the edge or arrow is a direct (relative to the features represented in the diagram) cause of the feature represented by the node at the head of an arrow or directed edge. Diagrams such as these are *directed graphs*, and carry with them obvious notions – a node at the head of an arrow is the *child* of a node at the tail, which is its *parent*; some nodes are *ancestors* of others, their *descendants*; there are *paths* from ancestors to descendants, and so on.

I will say a system *S* of variables is *causally sufficient* provided that for every pair of variables *X*, *Y* in *S*, if there is a variable *Z* from which there is a causal path to *X* and also a causal path to Y, and the two paths do not intersect except at *Z* then *Z* is in *S* as well. Informally, variables in a causally sufficient set have no 'latent' common causes.

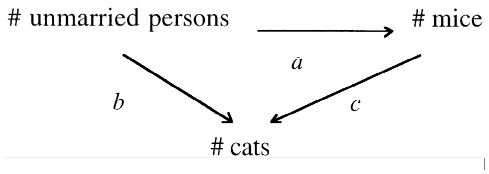
The independence claim I made about Baumrind's example – that it implies that number of unmarried persons and number of mice are independent given number of cats – assumed that intervening causes 'screen off' more remote causes from their effects. The directed graph formalism permits this assumption about the connection between causal structure and probabilities to be put much more generally. In Baumrind's example the nodes were variables that could take different numerical values in different villages. The village were the units, and there was an implicit relevant population of units.

> *Markov Condition*: In a causally sufficient system described by a directed acyclic graph, *G*, conditional on any set of values of all of its parents, every variable is independent (in probability) of the set of its non-descendant, non-parents in *G*.

In the directed graph for Baumrind's example, the parent set of number of mice is {number of cats} and the set of all non parent, non-descendants of number of mice is {number of unmarried persons} so the Markov condition gives exactly the conditional independence I asserted.

The Markov condition is not given by God, but it is not easily avoided either. It is a necessary feature of every causally sufficient system in which effects are (measurable) functions of their causes, and the unexplained variables (exogenous variables in economists' terms; independent variables in psychologists' terms; in graphical terms, the variables represented by nodes without edges into them in the directed acyclic graph) are jointly independent in probability. Almost every causal model I have ever come across in the psychological or social science literature conforms to it. In particular, all regression models, whether linear or logistic, with a causal interpretation, all factor analysis models, all 'recursive structural equation' models or path models,[2] time series models with a causal interpretation, and so on.

The three associations and the single conditional independence in Baumrind's example could have been explained in another way besides the five alternatives I offered. Those phenomena could have been explained by supposing (i) that number of unmarried persons has a direct influence on number of mice, having nothing to do with number of cats; (ii) that both number of unmarried persons and number of mice influence number of cats. Graphically:



How could this explain the conditional independence of unmarried persons and mice given cats? This way: The edge marked "a" creates an association between # unmarried persons and # mice, no matter whether # cats is only conditioned on. But the two edges that collide at # cats create an association between # unmarried persons and # mice only *conditional* on # of cats. So if the association marked by the a edge is *perfectly canceled* by the conditional association produced by the collision of the two edges marked b, c – if the two associations are equal but of opposite signs – then # unmarried persons an # mice will be independent conditional on # of cats. For example, if the relations are linear, and a, b, c represent linear coefficients, as in:

$$\#\text{ mice} = a\ (\#\text{unmarried persons}) + e1$$

$$\#\text{ cats} = b\ (\#\text{unmarried persons}) + c\ (\#\text{ mice}) + e2$$

where *e*1 and *e2* are independently distributed unobserved causes ('noises') and observed variables are standard normal (meaning they are normally distributed and scaled by convention to have zero means and variances of 1), then all of the requirements of the associations and conditional independence will be satisfied provided a is negative, b is positive, c is negative, and $a = -bc$.

In the ignorance I imagined, this explanation will save the phenomena, but it seems inordinately – even unscientifically – complex, and in the absence of prior knowledge I think most people would reject it in preference to explanations 3 or 4 above. Notice that the conditional independence does not result from the Markov condition applied to the causal graph just given – the Markov condition applied to that graph gives no independencies whatsoever. And that observation leads to a general formulation of the intuition about simplicity the example illustrates:

> *Faithfulness Condition*: For any variables X, Y and any set of variables Z in a causally sufficient system described by a directed acyclic graph G, X is independent of Y conditional on Z if and only if the Markov condition applied to G implies that X is independent of Y conditional on Z.

Faithfulness is easier to evade than the Markov condition, but not very easy. Both for linear stems and for systems of variables each having only a finite number of values, 'almost all' probability distributions that satisfy the Markov condition for a directed acyclic graph also satisfy the faithfulness condition.[3]

The Faithfulness assumption seems to bitterly divide scientists, even when they have not formulated it explicitly. The late, eminent sociologist, Hans Zeisel, resigned from the board of supervisors of an experiment funded by the Department of Labor when the principals of the experiment saved their favorite hypothesis by forwarding an unfaithful explanation of the data. The principals in turn mounted a rather vicious *ad hominem* attack on Zeisel. (See Glymour et al., 1987 for a review and references). In cognitive psychology, recent disputes over unconscious mechanisms of recall turn exactly on whether the faithfulness assumption is accepted or rejected. (See Jacoby et al., 1997 for a discussion and references).

The directed graph representation of causal mechanisms, with these two conditions, or with various modifications of the Faithfulness condition, provides a key to predicting the results of interventions, and thus to planning, and the basis for studying causal inference both in natural and artificial systems. Before developing those implications, we should pause to illustrate how the representation, if understood and used, might help avoid some mistakes in the interpretation of psychological experiments.

5. Revisiting a Psychological Experiment on Causal Judgement

In order to follow the published interpretation of the following experiment, reported by Baker et al. (1993), the reader must understand that a received view among many psychologists (Allan and Jenkins, 1983) who work on causal judgement is that the correct, normative, measure of the influence of a cause c on an effect e is the 'contingency' measured by:

$$\Delta P = \text{Prob}(e|c) = \text{Prob}(e| \sim c).$$

This will sound ludicrous to any philosopher, statistician or social scientist familiar with confounding, but, with exceptions to be noted later, the community of psychologists interested in causal judgement seems approximately partitioned into those who hold this opinion and those who hold the ‘mechanist’ view discussed previously.

Baker, et al. report the following experiment, designed to show that in the presence of causes with a large influence on an outcome variable, human subjects underestimate the influence of less important variables.

Subjects played computer games in which they tried to move a tank icon through a ‘minefield.’ Subjects had the power to camouflage or not camouflage the tank on each trial. Sometimes an airplane icon would appear. In the first experiment reported, the computer arranged things so that the following probabilities obtained, where *O* represents getting through the minefield, *C* is camouflage, *P* is appearance of the plane:

$P(O \mid P)$ = 1

$P(O \mid \sim P)$ = 0

$P(O \mid C)$ = 0.75

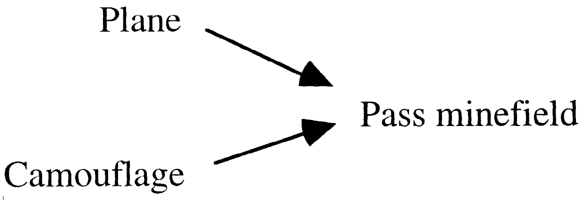
$P(O \mid \sim C)$= 0.25

Before, during and after each game, which consisted of 40 trials, subjects were asked to estimate, on a scale from -100 to +100, ‘the effectiveness of the camouflage’ (417).

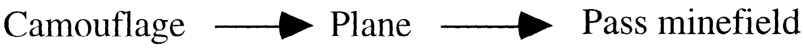
The authors write that ‘The crucial finding of this experiment concerns the effect of the high-plane contingency on the estimates of the effectiveness of the camouflage. A high plane contingency sharply reduced estimates of the effectiveness of both the positive (= 0.5) and the zero camouflage contingencies.’ (418). For example, in the experiment with the probabilities specifies above, the mean subject assessment of the ‘effectiveness’ of camouflage was 0.06 (that is, 6 on the -100 to +100 scale) rather than ΔP = 0.5 (50 on the -100 to +100 scale).

The problem that concerns me was pointed out in slightly different terms by Barbara Spellman (1996a). Consider the causal process, the mechanism, in the experiment. The real causal process was that, with the help of a randomizer, the subjects’ choice of camouflage/not-camouflage caused the tank icon to pass/not pass the minefield icons, and the intermediate data structure that produced passing/not passing the minefield icons also produced, deterministically, the appearance/absence of the plane. This set-up had consequences the authors did not note, in particular in the experiment, camouflage/not camouflage and plane appearance/absence are statistically dependent, and camouflage/not camouflage is independent of passing/not passing through the minefield *conditional* on appearance/absence of plane.

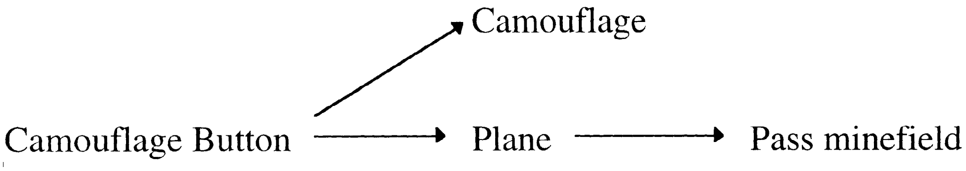
The causal process as it appeared to the subjects, initially, however, looked something like this:



The cover story and context made the subjects assume that the plane's appearance or absence was a cause, not an effect, of the tank passing or not passing the minefield. By 40 trials the subjects appear to have learned that camouflage state and passing/not passing the minefield are independent conditional on plane presence/absence. Subjects were not asked to give their picture of the causal relations, but the only causal pictures consistent with experimental cover story and context forced on the subject, and with the probabilities the trials exhibited, are



or more likely:



If the subjects had the second of these pictures, or if they understood 'effectiveness of camouflage' to mean something like, 'effectiveness controlling for other causes' and had the first of these pictures, then their mean answer was nearly optimal.

Unless psychological experimenters have the right picture of the causal mechanism and its connection with the frequencies that subjects observe, and a correct picture of how the problem of causal inference should appear to their subjects, results about human deviation from 'norms' are apt to be nonsense.

6. Parameters, Networks and Psychological Theories

As in Baumrind's example, the simple network representation behind models of causal mechanisms is usually hidden by equations, or parameters, or informal descriptions. Revealing that structure can illuminate the hypothesis and remove confusions, as in Baumrind's and Baker's examples but it can also sometimes provide generalizations and reveal important implications of a theory. An interesting illustration of the power of the graphical, or network, representation is afforded by applying it to Patricia Cheng's (1997) theory of human causal judgement.

Suppose *e* has a cause *i*, and let *a* represent all other causes of *e*. Assume *e* does not occur unless at least one of its causes occurs. Cheng reasons that the probability that *e* occurs given that *i* occurs is the probability that *i* causes *e* given that *i* occurs, plus the probability that *a* occurs given that *i* occurs times the probability that *a* causes *e* given that *a* occurs and *i* occurs, minus the probability that *a* occurs given that *i* occurs times the probability that both *a* and *i* cause *e* given that *a* and *i* both occur. She assumes the probability – for reasons that will be clear later, I denote it P(q_{ae}) – that a causes *e* given that *a* occurs is independent of whether *i* occurs, and likewise the probability, $P(q_{ie})$, that *i* causes *e* given that *i* occurs is independent of whether *a* occurs, and, further, that the probability $P(q_{aie})$ that *a* and *i* both cause *e* given that both occur equals $P(q_{ae})P(q_{ie})$. Hence she derives

$$\text{prob}(e = 1|i = 1) = \\ P(q_{ie}) + P(q_{ae})\text{prob}(a = 1|i = 1) - P(q_{ie})P(q_{ae})\text{prob}(a = 1|i = 1) \qquad (1)$$

which shows immediately that P(q_{ie}) is a conditional probability, specifically the probability that e occurs given that i occurs and a does not occur. *Cheng 's model of the power of a cause i to produce an effect e is, in this setting, the probability of e given that i occurs and that no other cause of e occurs.*

When *i* = 0

$$\text{prob}(e = 1|i = 0) = \\ P(q_{ae})\text{prob}(a = 1|i = 0) \qquad (2)$$

Now if *a* and *i* are independent, she deduces

$$\text{prob}(e = 1|i = 1) = P(q_{ie}) + P(q_{ae})\text{prob }(a = 1) - \\ P(q_{ie})P(q_{ae})\text{prob}(a = 1)$$

and

$$\text{prob}(e = 1|i = 0) = P(q_{ae})\text{prob}(a = 1)$$

Hence

$$\Delta P = \text{prob}(e = 1|i = 1) - \text{prob}(e = 1|i = 0) = \\ P(q_{ie}) - P(q_{ie})P(q_{ae})\text{prob}(a = 1) \qquad (3)$$

and finally

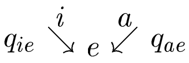
$$P(q_{ie}) = \frac{\Delta P}{1 - \text{Prob}(e = 1|i = 0)} \qquad (4)$$

Surprisingly, under these assumptions, if *a* and *i* are independent, the probability that *e* occurs given that *i* occurs and *a* does not can be estimated without observing the value of *a*. Cheng reviews a great deal of evidence that in problems in which subjects are asked to estimate the power of a facilitating (rather than inhibiting) cause *i*, and they are given reason to think i and all other causes *a*, of *e*, are independent, they estimate P(q_{ie}).[4]

Cheng's is a theory, one of the few I know of, that at least for special cases – binary variables and direct causes of an effect – addresses the subject's model of causal structure, the relations of that causal structure to probabilities, and the aim of judgements of causal power. And, more to the good, it does not require of subjects extraordinary computational powers, tacit or explicit. The aim it supposes has a natural justification: the probability that *e* = 1 given *i* = 1 and all other causes a are absent does not depend on the frequency of other causes *a*, and so does not depend on the base rate of *e*, and can be extrapolated across contexts in which base rates differ. That is a useful and interesting property. Cheng gives the following example: suppose in an otherwise homogeneous population the probability of lung cancer among smokers is 0.95, while the probability of lung cancer among non-smokers is 0.9. Then if ΔP were the measure of power of smoking to produce lung cancer, that value would be 0.05. on a scale that goes from 0 to 1. But $P(q_{\text{smoking,lungcancer}})$ = 0.5 on a similar scale. Which is the more informative quantity? Suppose you learn that everyone in the homogeneous population was regularly exposed to asbestos?

O.K. so what's the connection between directed acyclic graphs and Cheng's model, and why is it important? No directed graphs are mentioned in Cheng's papers, there is nothing of the Markov condition or Faithfulness or any of that. The connection is that Cheng's model turns out to be a kind of directed acyclic graphical model known in the computer science literature as a 'noisy *or* gate.[5] The importance of the connection is that once it is recognized, a little mathematics enables the formulation of a considerably more general theory – or, put another way, permits us to draw consequences and make computations for much more complex instances of her model.

In Cheng's model, *i* is a cause of *e* and *a* represents all other causes, assumed unobserved. So there is an obvious graph in her simplest model in which *i* and *a* are independent.



The noisy *or* gate is given by the equation

$$e = iq_{ie} \bigoplus aq_{ae} \tag{5}$$

where *e, i, a*, q_{ie}, q_{ae} are binary variables and the $\oplus$ is Boolean addition. Assume $\{$*i, a*, q_{ie}, $q_{ae}\}$ to be jointly independent. Then for the noisy *or* gate,

$$\begin{aligned} P(e=1) - P(iq_{ie} \bigoplus aq_{ae} = 1) \;=\; & P(i=1)P(q_{ie}=1) + P(a=1)P(q_{ae}=1) \\ & -P(a=1)P(q_{ae}=1)P(i=1)P(q_{ie}=1). \end{aligned}$$

The parameter $P(q_i = 1)$, for example can be estimated by

$$P(q_{ie} = 1) = P(e = 1|i = 1, a = 0)$$

or, when *a* is unobserved, by Cheng's formula,

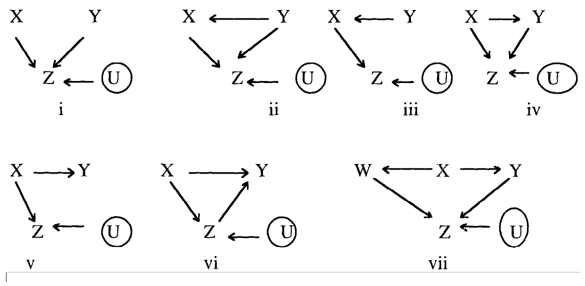
$$P(q_{ie}) = \frac{\Delta P}{1 - \text{Prob}(e = 1|i = 0)} \tag{6}$$

Cheng's model of human judgment of positive causation *just is* a directed acyclic graph parameterized as a noisy-or-gate.

So what?

7. Generalizations and Implications

Cheng and her collaborators, her former student, Spellman, and increasingly others as well, have emphasized the possibility that in problems with more than one potential cause, subjects judgements of causal efficacy focus on *conditional* contingencies, as in the Baker example. But, as Spellman and Cheng both note, to understand what subjects are doing, and to understand what they are doing right and what they are doing wrong, we must understand *which* conditional probabilities are people assessing when they answer questions about causal efficacy, and why they *ought* to assess some conditional probabilities rather than others. For example, consider judging the power of *X* to influence *Z* in the following alternative causal systems in which *U* represents unobserved causes of *Z*:



What does Cheng's model imply about how causal power is estimated in these cases? The graphical representation with noisy or-gate parametrization reveals ambiguities in the question, and permits a general solution.

Consider case (ii). The noisy or gate equations are:

$$\begin{aligned} Z &= Xq_{xz} + Yq_{vz} + Uq_{uz} \\ X &= Yq_{yx} \end{aligned}$$

The causal power of to produce Z *is the probability that* Z = 1 *given that* X = 1, Y = 0 and U = 0. But the second equation above implies that if Y = 0 then X = 0, so this probability is undefined. In case (ii), the parameter $P(q_{xz})$ is not this probability at all but rather the *probability that Z* = 1 *given that* Y = 0 and U = 0 *and given an intervention to bring about X* = 1. An intervention that forces the value 1 on *X*, regardless of the value of *Y*, transforms the causal structure of case (ii) into the causal structure of case (i). In case (i) the probability (of *Z* = 1, etc.) conditional on *X* = 1 is equal to the probability given an intervention to bring about *X* = 1, but in case ii the quantities are distinct. A general theory of the transformations of causal structures and probabilities under ideal interventions is given in Spirtes (1993) and Pearl (1995).

Cases iv and vii pose another difficulty. Cheng's measure of the causal power of X to produce Y – the probability that *Z* =1 given (an intervention to produce) *X* = 1 and *all* other causes of *Z* have value 0 – is necessarily zero in case vii, because *X* only influences *Z* through effects of *X* (namely *W* and *Y*), which in turn are causes of *Z*. That suggests that an alternative measure of the causal power of *X* to produce *Z* might be the probability that *Z* = 1 given that *X* = 1 and that all other causes of *X that are not effects of X* are 0. Call the original measure *direct* causal power and the new measure *total* causal power. In case iv, the two measures of causal power are distinct, but both non-zero. In these cases any simple request for a judgment of causal power is ambiguous. In the other cases the two measures are equal.

The noisy or gate representation permits us to estimate both the total and the direct causal power in a broad class of cases including these seven:

> *Proposition*: Consider any directed acyclic graph, parameterized as a noisy or gate, representing the mechanism involving *c* and *e*, and having no unobserved variable *u* with two non-intersecting paths from *u* respectively to *c* and *e*. Consider each directed path (each causal pathway) from *c* to *e*. Form the product of the q coefficients associated with the links on each path, then take the Boolean sum of these products over all paths from *c* to *e*. The probability of that Boolean sum is the total causal power of *c* to produce *e*, that is, the parameter whose value is the probability of *e* given an intervention to bring about *c*, and given that all other causes of *e*, that are not themselves effects of *c*, are absent. That probability is given by the generalization of Cheng's formula (4), using on the r.h.s. probabilities conditional on the absence of all observed causes of *e* that are not effects of *c*.

Cheng's original estimation formula, conditionalized, remains correct for total causal power, but then estimates an algebraic combination of noisy or gate parameters. The reader familiar with linear models will note the similarity between the rule for compounded noisy *or* gates, and the rules for calculating effects in path analysis.

Under the same assumptions, the direct causal power of *c* to produce e can be calculated as in the Proposition, but by deleting the phrase 'that are not effects of *c*' in the last sentence. The proposition may be clarified by computing the total causal power for an example:

The causes of *e* that are not effects of *c* are *f*, *h* and *g*. According to the proposition the total causal power of *c* to bring about *e* is:

$$P(e = 1|f = 0, g = 0, h = 0, c = 1 \text{ by intervention,}) = P(q_{cb}q_{be} \bigoplus q_{cd}q_{de} = 1)$$
$$= [P(e{=}1|c{=}1, h{=}0) - P(e{=}1|c{=}0, h{=}0)]/(1 - P(e{=}1|c{=}0, h{=}0)].$$

The r.h.s. of this equation is conditioned on *h*=0 because h is an observed cause of *e* but not a descendant of *e*. Recall that by the convention used in directed graphs, *f* and *g* are independent of *c*. Here is a derivation:

$$\begin{aligned} e &= q_g g \bigoplus q_b b \bigoplus q_d d \\ &= q_g g \bigoplus q_b(q_f f \bigoplus q_{cb}c) \bigoplus q_d q_h h \bigoplus q_d q_{cd} c \\ &= (q_g g \bigoplus q_b q_f f \bigoplus q_d q_h h) \bigoplus C(q_b q_{cb} \bigoplus q_d q_{cd}) \end{aligned}$$

When $c = 1$ by intervention, and $h = 0$:

$$e = (q_g g \bigoplus q_b q_f f) \bigoplus (q_b q_{cb} \bigoplus q_d q_{cd})$$

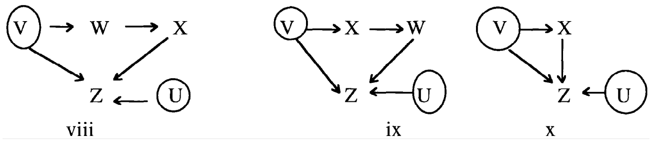
When $c = 0$ and $h = 0$:

$$e = q_g g \bigoplus q_b q_f f$$

Let $\Delta P_{h=0} = P(e = 1|c = 1 \text{ by intervention, } h = 0) - P(e = 1|c = 0, h = 0)$
Then

$$\begin{aligned} \Delta P_{h=0} &= P(q_b q_{cb} \bigoplus q_d q_{cd}) - P(q_g g \bigoplus q_b q_f f) \bullet P(q_b q_{cb} \bigoplus q_d q_{cd}) \\ &= P(q_b q_{cb} \bigoplus q_{cd} q_d)(1 - P(e = 1|c = 0, h = 0)]. \end{aligned}$$

Cases with unobserved causes of observed cause and effect are not covered by the proposition given above, but the graphical, noisy or gate representation shows that Cheng's theory has implications for them.



.

In case (x) the causal power of *X* to produce *Y* cannot be calculated; I do not know if it can be calculated in case (ix); in case viii it can be calculated and equals

$$P(Z = 1|X = 1, V = 0, U = 0) = P(qXZ) =$$
$$\frac{P(Z = 1|X = 1, W = 0) - P(Z = 1|X = 0, W = 0)}{1 - \text{Prob}(Z = 1|X = 0, W = 0)}$$

where '*X*=1' means 'by intervention.'

8. Still More General Generalizations

Cheng’s model assumes causes and effects have only two values: present or absent. But in some contexts causes may have no natural ‘absent’ value: consider height and weight of persons. When several such causes contribute to an effect, it may not make sense to ask for the power of a particular cause, *c*, to produce effect *e* when all other causes are absent, or even to ask for the causal power of a particular value of *c* to produce a particular value of *e* when all other causes are absent. In these settings, however, we can still consider *the average or expected probability of e (or a particular value of e) given an intervention to bring about c (or a particular value of c)*. This measure – call it ΔI – does not condition on the absence of other causes of *e* besides *c*, but instead averages over their values. Unlike Cheng’s, this measure is obviously not invariant over contexts in which the frequencies of other (besides *c*) causes of *e* vary, but in any context in which those frequencies are reasonably stable, the measure, when it can be estimated, gives a more accurate prediction of the results of an ideal intervention, and fits more closely with both intuition and scientific practice, than the ΔP so nearly ubiquitous is psychology.

There is a general theory about how to compute ΔI (Sprites, et al, 1993; Pearl, 1995). ΔI can be computed for each of the causal systems illustrated above, except for structure x. ΔI is not in general equal to ΔP conditional on any set of observed variables. For example, ΔI can be computed for structure ix but is not equal to any conditional contingency.

9. Unexamined Psychological Questions

There are distinct aspects to learning causes: on the one hand, learning the structure, or topology of the causal graph, what causes what; and, on the other hand, learning the parametrization associated with the causal structure. An intelligent system learning about the world must learn both, because the parametrizations have no sense without the causal topology. Of course the two might in fact be learned together, or aspects of either might be already known to the learner, who needs only a sufficient completion of the causal story, which may require more knowledge of the causal graph and more knowledge of the parameters. *A central question in cognitive psychology therefore ought to be how humans are able to learn both the causal structure and its parametrization, or aspects of the two together.*

That question has two particular versions:

(1) Given background knowledge about the causal structure or topology, how do people judge – and learn – the causal power or efficacy of an indirect cause – direct or total – which may be confounded with the effect by various common causes, and which may influence the effect through several mechanisms or pathways in the causal graph?

(2) Given limited knowledge about the causal structure or topology, how do people use observations of associations to expand that knowledge?

As to the first of these questions, there are experiments that support the hypothesis that people estimate causal power in accord with Cheng's model in cases (i) and (ii) above, but what happens (when the aim of judgement is disambiguated between direct and total causal power) in cases such as (iv) through (ix) does not seem to be known.

The second question has been addressed, but in a remarkably limited way. The psychological literature focuses on experiments that provide the subject with sufficient context – sufficient prior knowledge about causal structure, that only the value of a single parameter remains to be learned. Typically, some special value of that parameter corresponds to the absence of a single edge in the causal graph, an edge whose direction, if it exists, is already known to the subject. This may in fact be the only way *humans* extend their knowledge of causal structure, or it may not. It is certainly not the only possible way.

Consider the following example: data are available on the associations of four variables *X, Y, Z, W*. Nothing is otherwise known about the time order or causal relations among these variables, except that the values of these variables for a unit did not influence whether the unit was sampled. Suppose the associations show the following pattern:

X and Y are independent

X and Y are independent of W conditional on Z

No other independencies hold

Then, under the assumptions previously discussed, it follows necessarily that *Z* causes *W*. That is, *every* directed graph (and probability distribution), together satisfying the Markov and Faithfulness conditions, and which implies the three features just listed, contains a directed edge

$Z- > W.$

It doesn't matter whether the graph does or does not contain unobserved common causes of any pair of *X, Y, Z, W*, it must contain the edge from *Z* to *W*.

Or consider the following example: data are available on the associations of four variables *X*, *Y*, *Z*, *W*. Nothing is otherwise known about the time order or causal relations among these variables, except that for no unit did the values of these variables influence whether the unit was sampled. Suppose the associations show the following pattern:

X is independent of $\{Z,W\}$

W is independent of $\{X,Y\}$

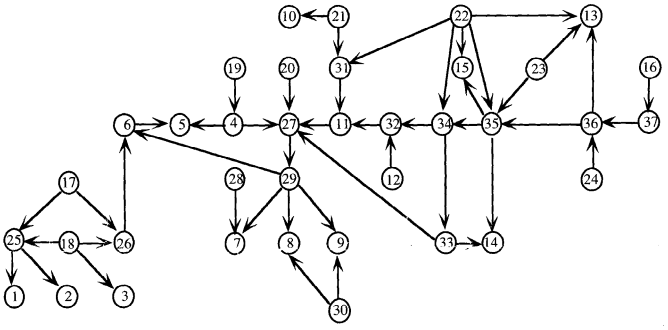
No other independencies hold

Then, under the assumptions previously discussed, it follows necessarily that there is a common cause of *Y* and *Z* other than *X* or *W*. That is, *every* directed graph and probability distribution, together satisfying the Markov and Faithfulness conditions, which implies the three features just listed contains another vertex, call it *U*, and a pair of directed paths from *U* to *Y* and to *Z*.

These simple examples illustrate that associations alone, under appropriate assumptions, sometimes suffice to determine causal connection, the direction of causation, and even the presence of unobserved or unnoticed causes. With stronger assumptions about prior knowledge, still more can be learned from the patterns of independencies and dependencies. For example, if it is known that there are no unobserved common causes of measured variables, then everything about the causal graph can be learned, except that any two graphs that have all of the same ‘unshielded collider’ structures – for example,

$$X \rightarrow Y \leftarrow Z$$

will be indistinguishable. For example, without the use of any prior causal information, computer algorithms are able to infer from associations most of a model of an emergency medical system show below (taken from Beinlich et al., 1989)



KEY:

1 – central venous pressure
2 – pulmonary capillary wedge pressure
3 – history of left ventricular failure
4 – total peripheral resistance
5 – blood pressure
6 – cardiac output
7 – heart rate obtained from blood pressure
8 – heart rate obtained from electrocardiogram
9 – heart rate obtained from oximeter
10 – pulmonary artery pressure
11 – arterial-blood oxygen saturation
12 – fraction of oxygen in inspired gas
13 – ventilation pressure
14 – carbon-dioxide content of expired gas
15 – minute volume, measured
16 – minute volume, calculated
17 – hypovolemia
18 – left-ventricular failure
19 – anaphylaxis
20 – insufficient anesthesia or analgesia
21 – pulmonary embolus
22 – intubation status
23 – kinked ventilation tube
24 – disconnected ventilatio tube
25 – left-ventricular end-diastolic volume
26 – stroke volume monitor
27 – catecholamine level
28 – error in heart rate reading due to low cardiac output
29 – true heart rate
30 – error in heart rate reading due to electrocautery device
31 – shunt
32 – pulmonary-artery oxygen saturation
33 – arterial carbon-dioxide content
34 – alveolar ventilation
35 – pulmonary ventilation
36 – ventilation measured at endotracheal tube
37 – minute ventilation measured at the ventilator

The causal graph encodes the conditional independencies that the experts specified among these variables; the search algorithms find the graphs that explain the resulting patterns in the data.

There are a number of algorithms in the computer science literature, using a variety of techniques, that recover causal structure from associations by taking advantage of these relationships between causal structure and patterns of constraints on association. The computational complexity of the procedures depends on the complexity of connections in the causal graph generating the data – strictly, on how many parents each variable has, on average. For sparse graphs the procedures are very fast, and with good data can recover a great deal of structure quite reliably. Whether humans can do the same, at least in simple cases, is essentially unknown. Only one experiment has been reported. Hashem and Cooper, 1996, gave medical students information about associations for a variety of cases involving two and three binary variables, described as disease or gender features. The subjects were asked, essentially, to recover the causal graph from the associations, and their responses were compared with those of a Bayesian inference algorithm using the same associations. The subjects did poorly in problems with three variables, but the result is uncertain for two reasons. First, because of sample size and the exclusive use of binary variables, spurious near-independencies held in the data given the subjects, so that in important cases the Bayesian search algorithm performed comparably poorly. Second, a lot of experience in psychological experiments suggests that humans do much better with frequency and independence judgements when they are not given numbers, but instead actually observe the events or can see the frequencies displayed graphically, or both.

Besides association and short of experimental intervention, the cue to causation most commonly available to us is the order of occurrence of events. Causes do not come after effects. Knowledge of time order considerably speeds up algorithmic search for structure and increases the reliability of output as well. Perhaps more important for understanding human inference, knowledge of time order may compensate for circumstances in our environment in which the Faithfulness assumption does not hold. It may be that for many of the causal relations of everyday life, relationships among observed features or variables are nearly deterministic. Faithfulness fails to hold in many deterministic systems, for technical reasons I will pass by. But when time order is known, the Faithfulness assumption is not needed for inference to causal structure in systems of deterministically related observed variables; in those contexts Faithfulness can be replaced by the weaker assumption that multiple mechanisms relating two variables do not perfectly cancel one another.

10. Conclusion

For computers anyway, there is an answer to the puzzle posed by Shanks’ claim that causes are learned from associations. The answer–algorithms that represent causal structure by networks or directed graphs and infer aspects of that structure from data about frequencies, relying on very broad assumptions connecting causal structure with conditional probabilities–raises a host of unexamined issues for experimental psychology. The same representation also provides a mathematical tool for generalizing and analyzing leading theories of human Judgement of causal power, provides the means to see and articulate important distinctions about causal influence – distinctions that, if not recognized, easily lead to erroneous interpretations of psychological experiments – and provides a coherent norm against which to measure human judgement.

Notes

[1] I thank Patricia Cheng for several months of illuminating conversation on the subject of this paper, and Alison Gopnik for bringing Cheng's work to my attention. Section 2 of this paper borrows from Glymour and Cheng, in press. Most of this paper was written while I was a Fellow of the Center for Advanced Study in the Behavioral Sciences, supported by a grant from the Andrew Mellon Foundation.

[2] Without correlated errors, which typically can be treated as the effect of unobserved latents, following Simon. Work in the last two years by Peter Spirtes and Thomas Richardson has shown that a generalization of the Markov condition, due to Judea Pearl, also applies to 'non-recursive' structural equation models.

[3] Further, if (i) causation is transitive (and of course irreflexive), (ii) for every value y of an effect variable *Y*, there is some set of values of the causes of *Y* that determine the value y, and (iii) if *U* is a parent of *X*, then for any set of values of all other parents of *X*, *X* is a non-constant function of *U*, and (iv) every variable in a causally sufficient set *S* has a cause not in *S*, and the probability measure is strictly positive, then the probability measure is faithful to the graph for *S*.

[4] Cheng provides an analogous, equally well motivated, analysis of the power of inhibiting causes and evidence that subjects in appropriate circumstances estimate it, but I will not explore it here.

[5] The connection was pointed out to me by Peter Spirtes.

References

Ahn, W. and Bailenson, J. (1996), 'Causal Attribution as a Search for Underlying Mechanisms: An Explanation of the Conjunction Fallacy and the Discounting Principle', *Cognitive Psychology*.

Ahn, W., Kalish, C.W., Medin, D.L. and Gelman, S.A. (1995), 'The Role of Covariation Versus Mechanism Information in Causal Attribution', *Cognition* 54, pp. 299–352.

Allan, L.G. (1980), 'A Note on Measurements of Contingency Between Two Binary Variables in Judgment Tasks'. *Bulletin of the Psychonomic Society* 15, pp. 147–149.

Allan, L.G. and Jenkins, H.M. (1983), 'The Effect of Representations of Binary Variables on Judgment of Influence'. *Learning and Motivation*, 14, pp. 381–405.

Anderson, J.R. and Sheu, C-F. (1995), 'Causal Inferences as Perceptual Judgments'. *Memory & Cognition* 23, pp. 510–524.

Baker, A.G., Mercier, P., Valle-Tourangeau, F., Frank, R. and Pan, M. (1993), 'Selective Associations and Causality Judgments: The Presence of a Strong Causal Factor May Reduce Judgments of a Weaker One', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, pp. 414–432.

Cheng, P.W. (1997), 'From Covariation to Causation: A Causal Power Theory'. *Psychological Review* 104, pp. 367–405.

Cheng, P.W. and Novick, L.R. (1992), 'Covariation in Natural Causal Induction'. *Psychological Review*, 99, pp. 365–382.

Glymour, C. and Cheng, P. W. Causal Mechanism and Probability: A Normative Approach', in M. Oaksford and N. Chater (eds.), *Rational Models of Cognition*. Oxford, U.K.: Oxford University Press (in press).

Harré, R. and Madden, E.H. (1975), *Causal Powers: A Theory of Natural Necessity*, Totowa, New Jersey: Rowman & Littlefield.

Hashem, A.I. and Cooper, G.F. (1996), 'Human Causal Discovery From Observational Data'. *Proceedings of the 1996 symposium of the American Medical Information Association*.

Jacoby, L., Yonelinas, A. and Jennings, J., (1997), The Relation Between Conscious and Unconscious (Automatic) Influences: A Declaration of Independence, in J. Cohen and J. Schooler (eds.), *Scientific Approaches to Consciousness*., Mahwah, N.J., Lawrence Erlbaum Associates, pp. 13–47.

Jenkins, H. and Ward, W. (1965), 'Judgment of Contingency Between Responses and Outcomes'. *Psychological Monographs* 7, pp. 1–17.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann.

Pearl, J. (1995), 'Causal Diagrams for Empirical Research', *Biomtrika* 82(4), pp. 669–709.

Price, P.C. and Yates, J.F. (1993), 'Judgmental Overshadowing: Further Evidence of Cue Interaction in Contingency Judgment'. *Memory & Cognition* 21, pp. 561–572.

Rescorla, R.A. (1968), 'Probability of Shock in the Presence and Absence of CS in Fear Conditioning'. *Journal of Comparative and Physiological Psychology* 66, pp. 1–5.

Shanks, D.R. (1995), 'Is human learning rational?' *Quarterly Journal of Experimental Psychology*, 48A, pp. 257–279.

Shultz, T.R. (1982), 'Rules of Causal Attribution'. *Monographs of the Society for Research in Child Development*, 47, (1).

Spellman, B.A. (1996a), 'Acting as Intuitive Scientists: Contingency Judgments are Made While Controlling for Alternative Potential Causes'. *Psychological Science*, 7, pp. 337–342.

Spellman, B.A. (1996b), Conditionalizing causality, in D.R. Shanks, K.J. Holyoak, D.L. Medin (eds.), *The Psychology of Learning and Motivation*, vol 34: Causal learning (pp. 167–207). San Diego: Academic Press.

Spirtes, P., Glymour, C. and Scheines, R. (1993), *Causation, Prediction and Search*, New York: Springer.

Turner, M. (1987), *Death is the Mother of Beauty: Mind, Metaphor, Criticism*. Chicago: University of Chicago Press.

Waldmann, M.R. and Holyoak, K.J. (1992), 'Predictive and Diagnostic Learning Within Causal Models: Asymmetries in Cue Competition'. *Journal of Experimental Psychology: General* 121, pp. 222–236.

White, P.A. (1989), 'A Theory of Causal Processing', *British Journal of Psychology*, 80, pp. 431–454.

White, P.A. (1995), 'Use of Prior Beliefs in the Assignment of Causal Roles: Causal Powers Versus Regularity-based Accounts'. *Memory & Cognition*, 23, pp. 243–254.