1 Explaining Four Psychological Asymmetries in Causal Reasoning: Implications of Causal Assumptions for Coherence

Patricia W. Cheng, Laura R. Novick, Mimi Liljeholm, and Christina Ford

In this chapter, we describe four psychological phenomena that offer clues to how untutored people infer causal relations. We contrast the predictions for these phenomena, all involving asymmetries in causal inferences, according to two psychological approaches—an associationist approach (e.g., Cheng and Novick 1992; Jenkins and Ward 1965; Pearce 1987; Rescorla and Wagner 1972; Van Hamme and Wasserman 1994), and a causal approach (e.g., Cheng 1997, 2000; Novick and Cheng 2004). Our analysis reveals that each phenomenon is inexplicable by associationist models but follows coherently from a causal theory. What distinguishes these approaches is that the causal theory has the goal of explaining the occurrence of a target event by the potentially independent influences of candidate causes and other (background) causes. This goal has no analogue in associationist models. To arrive at a coherent explanation, the causal account creates a theoretical construct of causal power (Cartwright 1989)—the probability with which a cause influences an effect. According to this account, reasoners search for, or define, candidate causes with the goal of arriving at causes that influence a target effect independently of the background causes. In other words, they seek causes whose powers are (ideally) invariant regardless of how frequently the background causes occur (see Woodward 2003, for a discussion of the degree of invariance and depth of explanation; also see Haavelmo 1944, for a discussion of causes varying on degree of autonomy).

By ''cause,'' we mean both *simple* causes that consist of a single element and *conjunctive* causes that consist of a combination of two or more elements acting in concert; we also mean a direct cause in the sense that, for the purpose of analysis, intermediate causes that lie on the path between the candidate cause and the effect are ignored or treated as part of the

candidate. In our view, causal explanation occurs within a hypothesis-testing framework in which predictions based on various sets of assumptions are evaluated to reach the goal of a satisfactory explanation. This testing begins with simpler hypotheses unless there is evidence refuting them. Any processing system that cannot simultaneously evaluate all possible hypotheses needs an ordering bias; two reasons supporting a simplicity bias are: (1) simple causes are an inherent part of the definition of conjunctive causes (Novick and Cheng 2004), and (2) they are the elements in more complex networks. We restrict our discussion to causes and effects that are represented by binary variables with a present value and an absent value; this type of cause and effect, compared with the type represented by continuous variables, more clearly reveals the function of causal constructs. As will become clear, for situations that are well represented by binary variables of this type, associationist accounts do not allow for the possibility that the occurrence of the effect is the result of the independent influences of causes.

Psychology is not the only discipline that has inherited associationism. Some commonly used normative statistical measures—for example, the chi-square test and the cross-product ratio—are also associationist. Given that the asymmetry phenomena are manifestations of a coherent explanation of the occurrence of a target effect when causal assumptions are made, but only when such assumptions are made, a question arises: do these phenomena point to a basic problem that permeates both psychological and normative associationist models? When applied to test causal hypotheses, are associationist measures coherent? Before presenting the four phenomena, we first give a brief account of the two psychological approaches.

Alternative Psychological Accounts of Causal Learning

An Associationist Model

For brevity, we illustrate the pitfall of associationism primarily using only one such model, the ΔP model, which was independently proposed in philosophy (Salmon 1965) and in psychology (e.g., Jenkins and Wards 1965; Cheng and Novick 1992). It is the dominant associationist model of causal learning in the psychological literature; it makes the same predictions as Rescorla and Wagner's (1972) model—the dominant connectionist model of conditioning and causal learning—at equilibria if the two parameter

values of the effect in question in the latter model are assumed to be equally salient (e.g., Danks 2003). Other associationist models (e.g., Pearce 1987; Rescorla and Wagner 1972; Van Hamme and Wasserman 1994), despite their greater flexibility due to added parameters, are nonetheless unable to account for some robust psychological findings, including those to be discussed here (for more extended evaluations of these models, see Buehner, Cheng, and Clifford 2003; Cheng, Park, Yarlas, and Holyoak 1996; and Novick and Cheng 2004). The only exception is if Rescorla and Wagner's model is modified to become consistent with the explanatory causal construct just mentioned (see the noisy-OR and noisy-AND-NOT modifications in Danks, Griffiths, and Tenenbaum 2003).

According to the ΔP model, the *strength* of a causal relation between candidate cause i and effect e is estimated by the difference in the probability of e given i and given not-i:

$$\Delta P_i = P(e \mid i) - P(e \mid \bar{i}). \tag{1}$$

If ΔP_i is noticeably positive, reasoners are predicted to conclude that i causes e; if ΔP_i is noticeably negative, they are predicted to conclude that i prevents e; and if ΔP_i is not noticeably different from 0, they are predicted to conclude that i has no influence on e. We interpret equation 1 under the assumption that alternative causes are "controlled" (i.e., held constant); this interpretation is sometimes called the *probabilistic contrast* model (Cheng and Novick 1992).

The Causal Power Theory of the Probabilistic Contrast Model

The causal power theory of the probabilistic contrast model (Cheng 1997, 2000; Novick and Cheng 2004) instantiates an approach to causal learning that adds an explanatory layer to associationist models. *Causal power* is the theoretical probability with which a cause influences an effect e when the cause is present (Cartwright 1989); this influence can be *generative* (i.e., the cause produces e with a certain probability) or *preventive* (i.e., the cause prevents e with a certain probability). According to this causal power theory (Cheng 1997), to evaluate the power of a candidate cause i to influence e, reasoners partition all causes of e into candidate i and the composite of (known and unknown) causes of e alternative to i, labeled a here, and they explain covariation defined in terms of observable frequencies by the unobservable hypothetical causal powers of i and a. The partitioning of all

causes of e into i and a is the simplest possible conception of the causes of e that would still explain covariation between i and e.

The assumptions underlying the derivation of simple causal power are:

1. i and a influence e independently,
2. causes in a could produce e but not prevent it,
3. the causal powers of i and a are independent of their occurrences (e.g., the probability of a both occurring and producing e is the product of the probability of a occurring and the power of a), and
4. e does not occur unless it is caused.

Assumptions 1 and 2 are merely ''working hypotheses'' for the reasoner; they are adopted until the reasoner perceives evidence against them. If they are dropped, they are replaced by weaker assumptions, and alternative models apply (e.g., Cheng 2000; Novick and Cheng 2004).

With respect to assumption 1, *independent causal influence* is defined as follows: let us consider a simple case in which i and j, two causes of an effect e, are both present (j is different from a in that a, which includes unknown or unobserved causes, typically cannot be constrained to be present). If the *generative causal power* of i with respect to e is q_i (i.e., i produces e with probability q_i when i is present), and the generative causal power of j with respect to e is q_j, then if i and j influence e independently, the probability that e is produced both by i and by j would be $q_i \cdot q_j$. Under the same assumptions, the probability that e is produced by i *or* by j would be $q_i + q_j - q_i \cdot q_j$. This function relating the probability of e occurring (in the presence of i and j in this simple case) to the theoretical probabilities of e due to each constituent cause, is sometimes referred to as a noisy-OR gate (Glymour 2001).

When $\Delta P_i \geq 0$, reasoners evaluate the hypothesis that i produces e and estimate q_i. To do so, they allow the possibility that i produces e, and explain $P(e \mid i)$ by the probability of the *union* of two events: (1) e produced by i, and (2) e produced by a if a occurs in the presence of i. That is, they reason that when i is present, e can be produced by i *or* by a if a occurs in the presence of i. Likewise, they explain $P(e \mid \bar{i})$ by how often e is produced by a alone when a occurs in the absence of i.

Figure 1.1 illustrates these explanations of the two conditional probabilities by Euler circles. The dashed and undashed circles in the figure, respectively representing *e produced by i* and *e produced by a*, are both *unobservable*;

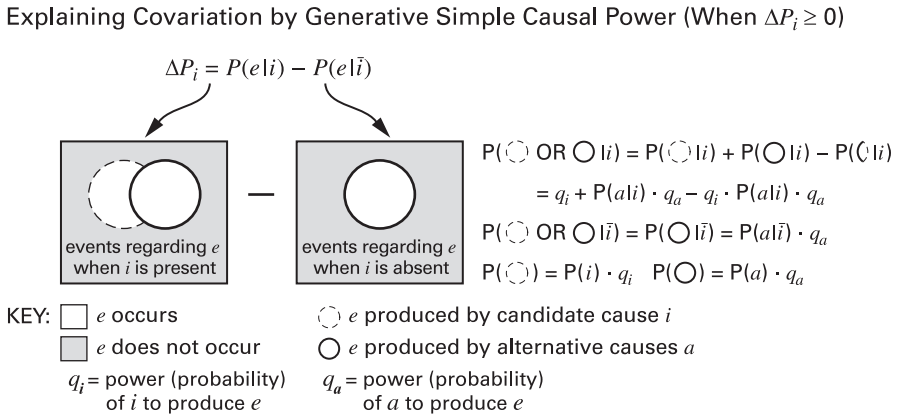

Figure 1.1
Euler diagrams illustrating the explanation of covariation in Cheng's (1997) theory of simple causal power (from Novick and Cheng 2004).

they are theoretical constructs. What can be observed is the shading—the white and shaded areas respectively representing *e occurring* and *e not occurring*; and the two boxes—respectively representing *exposure to i* and *no exposure to i*. Now, *e occurring* in the presence of *i*—the white area in the left box representing the *union* of the dashed and undashed circles—can be decomposed theoretically into the sum of the area of the dashed circle and that of the undashed circle, minus their overlap (see top equation in the figure). The relative size of the dashed circle in the left box (i.e., how often *e* is produced by *i* when *i* is present) depends on how often *i* occurs (always in this situation) and its causal power, q_i (i.e., how often *i* produces *e* when *i* is present). The undashed circle can be similarly explained. (See bottom two equations in the figure.) Likewise, the white area in the right box (when *i* is absent) can be analogously explained, in this case by how often *e* is produced by *a* in the absence of *i* (see middle equation). These explanations of *e* set up equations relating the observable quantities to the various theoretical variables.

For situations in which $\Delta P_i \leq 0$, there are analogous explanations for evaluating preventive causal power. The only difference in the preventive case is that reasoners evaluate the hypothesis that *i prevents e* rather than produces it. This difference implies that when reasoners evaluate whether *i* prevents *e*, they explain $P(e \mid i)$ by the probability of the *intersection* of two

events: (1) e produced by a if a occurs in the presence of i, and (2) e not stopped by i. That is, they reason that when i is present, e occurs only if it is both produced by a *and* not prevented by i. This function is sometimes referred to as a noisy-AND-NOT gate (Danks et al. 2003). The explanation of $P(e \mid \bar{i})$ remains as before.

The goal of these explanations of $P(e \mid i)$ and $P(e \mid \bar{i})$ is to yield an estimate of the (generative or preventive) power of i from observable frequencies alone, even though it may be impossible to observe the influence of i in isolation. For figure 1.1, this goal corresponds to estimating the size of the whole of the (invisible) dashed circle relative to the size of the left box (when i is present). These explanations show that, under some conditions but not others, covariation implies causation. One of the necessary conditions is "no confounding"—the independent occurrence of a and i. This condition is necessary when causal inference makes use of a probabilistic contrast, as we explain later.[1] In the figure, this condition corresponds to requiring that the undashed circles in the two boxes have the same size relative to their respective boxes. Whether one assumes that this condition is satisfied depends on perceived evidence for or against it.

When a and i do *not* occur independently (i.e., when there is confounding), one equation with four unknowns results, and there is no unique solution for q_i. But, in the special case in which a occurs independently of i (i.e., when there is no confounding), these explanations yield equations with only one unknown, the causal power of the candidate. Equation 2 gives an estimate of q_i when $\Delta P_i \geq 0$:

$$q_i = \frac{\Delta P_i}{1 - P(e \mid \bar{i})}; \tag{2}$$

and equation 3 gives an estimate of p_i, the *preventive* simple power of i, when $\Delta P_i \leq 0$:

$$p_i = \frac{-\Delta P_i}{P(e \mid \bar{i})}. \tag{3}$$

The two equations are logically related in that replacing e with not-e in one equation will yield the right-hand-side (RHS) of the other equation. That is, generating e is equivalent to preventing not-e.

Note that the RHS's of equations 2 and 3 require observations regarding i and e only (intermediate terms involving a drop out under no-confounding), implying that q_i and p_i can be estimated *without* observing

a. Also note that when $\Delta P_i = 0$, both generative and preventive powers can be evaluated; otherwise, the type of power to evaluate depends on the observed sign of ΔP_i. Finally, note that causal power has a well-defined meaning in terms of frequencies of events in the world (see Novick and Cheng 2004, for a contrast with associationist models in this regard). For example, $q_i = 1$ means that i is estimated to produce e in every entity, and $q_i = 0$ means that i is estimated to never produce e in any entity (i.e., to be noncausal). The values of p_i have analogous interpretations in terms of preventing e.

Myriad findings on psychological causal judgments that are inexplicable by associationist accounts are explained by this causal theory (see Buehner, Cheng, and Clifford 2003; Cheng 1997; Lien and Cheng 2000; Novick and Cheng 2004; White 2004; Wu and Cheng 1999). We discuss four such findings in the following sections.

Phenomenon 1: An Asymmetry between Cause and Effect

Empirical Finding

This asymmetry is revealed in a comparison across three figures, all taken from a psychological experiment on judgments of causal strength conducted on college students (Buehner, Cheng, and Clifford 2003). First, consider the pattern of information in figure 1.2. For this and similar figures, the subjects were told that the figure depicts the outcome of a fictitious experiment testing a medicine on allergy patients. The medicine was said to be effective for relieving the symptoms of allergies, but its unintended effects were undocumented; the study was conducted to evaluate one such possible side effect—headache. To encourage the no-confounding assumption, the patients were said to be randomly assigned to two groups—a group that did not receive the medicine (the top panel, the *control* group), and a group that did (the bottom panel, the *experimental* group). Each face in the figure represents a patient. A frowning face represents that the patient has a headache; a smiling face represents that the patient does not have a headache.

The subjects were asked to judge, based on the data in the figure, whether each medicine causes, prevents, or has no influence on headaches. If the medicine was judged to have an influence, a follow-up question asked them to rate the strength of the influence. For example, if subjects indicated that

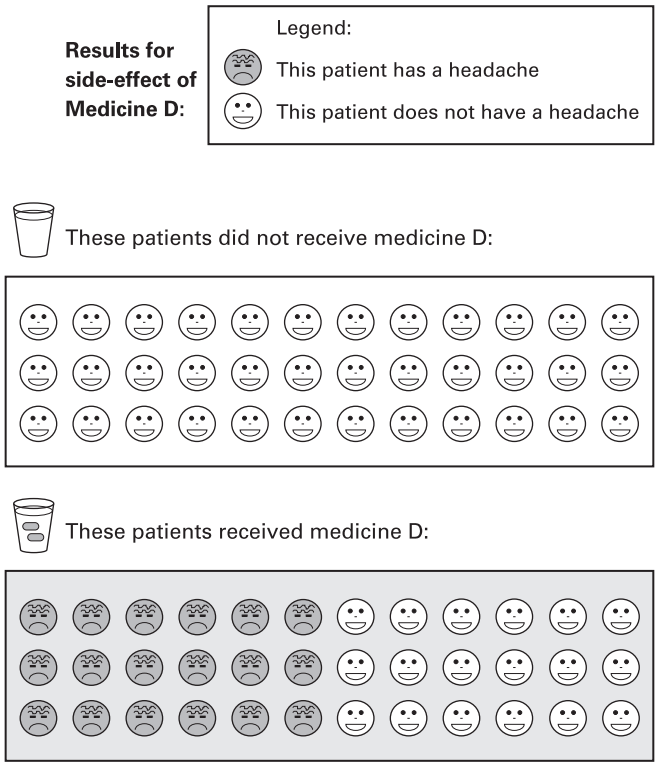

Figure 1.2
Stimulus material for Condition D from Buehner, Cheng, and Clifford (2003, Experiment 2) depicting the results of an experiment testing the side effect of a medicine.

a medicine causes headaches, they were asked to estimate, out of 100 allergy patients (selected from the same pool) who do not have headaches, the number who would get a headache if the medicine were given to them. Likewise, if subjects indicated that a medicine prevents headaches, they were asked to estimate, out of 100 allergy patients (from the same pool) all of whom have headaches, the number who would no longer have a headache if the medicine were given to them. A presentation of the information about individual patients in verbal, instead of pictorial form; patient-by-patient, in a sequence, instead of simultaneously as in our figures; produced the same pattern of psychological findings that we illustrate here.

For figure 1.2, most subjects indicated that medicine D causes headaches. They further estimated that if the medicine had been given to 100 allergy

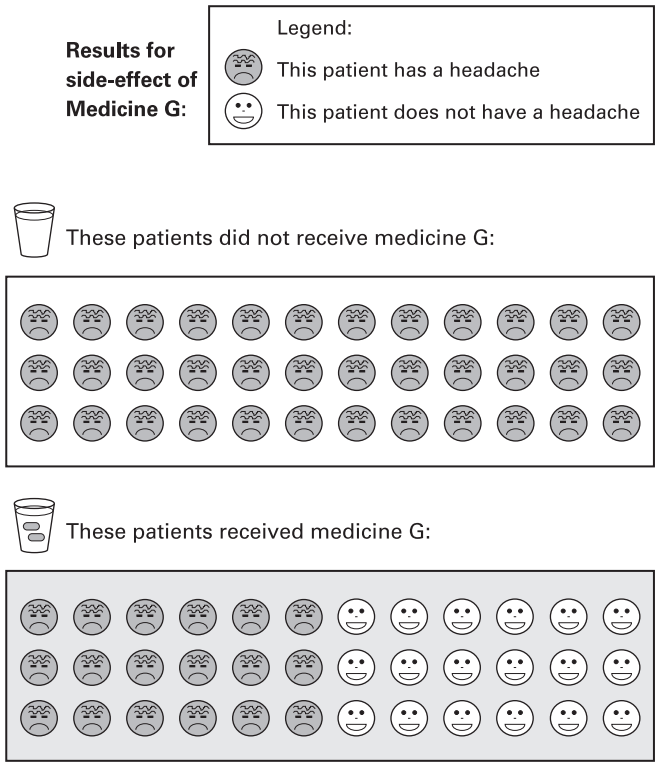

Figure 1.3
Stimulus material for Condition G from Buehner, Cheng, and Clifford (2003, Experiment 2).

patients who did not have headaches, 50 would get a headache from the medicine.

Next, consider the evaluation of the medicine in figure 1.3. The outcome pattern depicted in this figure results from changing each frowning patient in figure 1.2 into a smiling one, and vice versa (and grouping the resulting frowning patients and shifting them to the left). In other words, figure 1.3 transposes the two values of the outcome in figure 1.2. For this figure, most subjects answered that medicine G *prevents* headaches and estimated that about 50 out of 100 patients with headaches would no longer have headaches if given the medicine. Thus, from figure 1.2 to figure 1.3, there is a change in *causal direction*—from producing to preventing headaches—but no change in the magnitude of the strength of the causal relation.

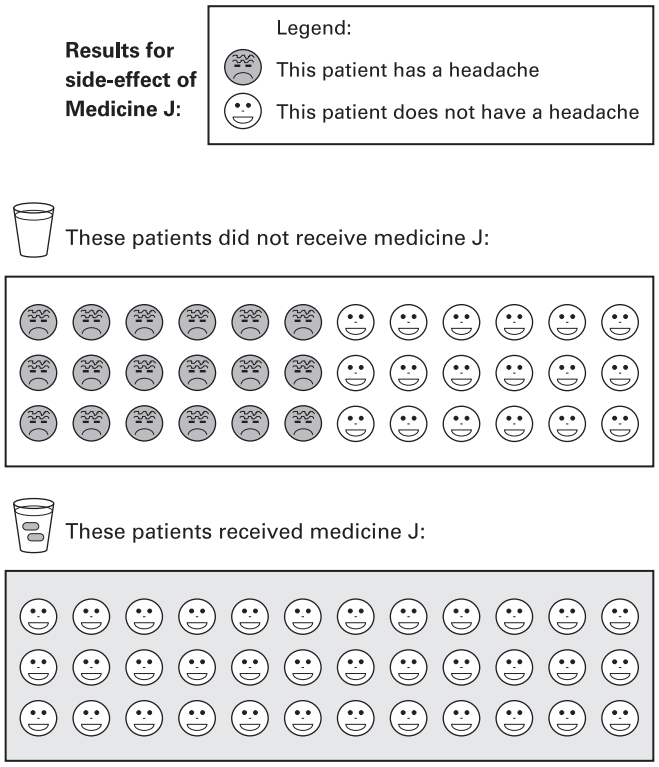

Figure 1.4
Stimulus material for Condition J from Buehner, Cheng, and Clifford (2003, Experiment 2).

Finally, consider figure 1.4. The outcome pattern in this figure is identical to that in figure 1.2 except that now, for every patient, the two values of the candidate cause are transposed; in effect, the top and bottom panels are switched. For this figure, most subjects responded that medicine J prevents headaches and estimated that it would do so in *all* 100 of the patients given the medicine. Note that in contrast to figure 1.3, the transposition of values doubled the estimated causal strength, from 0.5 to 1. These three figures illustrate that transposing the values of the effect (figure 1.3 vs. figure 1.2) has a consequence different from that of transposing the values of the candidate cause (figure 1.4 vs. figure 1.2).

Associationist models are unable to explain the different consequences of the two transpositions. Although the ΔP model correctly predicts a change in causal direction for each transposition, it gives the same absolute magni-

tude of strength, namely, 0.5, for all three figures. As an examination of equation 1 will show, the two variables in the equation (*i and e*) behave the same way when their values are transposed: for both variables, the two values are symmetrical in the sense that transposing them yields a ΔP that has the opposite sign but the same absolute value. Similarly, the commonly used chi-square test (e.g., Feinberg 1980) would yield the same χ^2 value when given data sets that are transformations of each other in the two ways just described. This test neither distinguishes between cause and effect variables, nor between the values of each variable; the values of each variable therefore have an identical status, as do the variables themselves, resulting in the symmetric treatment of all three figures.

Let us briefly illustrate how an associationist model with added parameters still fails to explain the pattern of results. For example, one could generalize equation 1 by adding weights to the conditional probabilities (e.g., Lober and Shanks 2000), so that weighted-$\Delta P = w_i \cdot P(e \mid i) - w_2 \cdot P(e \mid \bar{i})$. To fit the modal judgments for figures 1.2 and 1.4 (corresponding to an estimated weighted-ΔP of 0.5 for figure 1.2, and -1 for figure 1.4), the only solution is to set $w_1 = 1$ and $w_2 = 2$. This pair of weights, however, would predict a strength of -1.5 for figure 1.3, thus failing to explain the observed constant magnitude of causal strength across figures 1.2 and 1.3. As should be clear, post-hoc values for w_1 and w_2 that can explain the judgments for any two of the three figures will inevitably fail to explain the judgment for the third.

In contrast, the observed difference between the two transformations of figure 1.2 is explained by the causal power theory of the probabilistic contrast model (Cheng 1997, 2000). Recall that by the definition of causal power, a cause exerts its influence when it is present; it does nothing when it is absent. It therefore should be not be surprising that transposing the present and absent values of the candidate cause would yield different estimated causal strengths. For figure 1.2, instantiating equation 2 yields $q_i = 0.5$; for figures 1.3 and 1.4, instantiating equation 3 yields $p_i = 0.5$ and $p_i = 1$, respectively. These predictions are parameter-free.

An Incoherent Definition of Independent Causal Influence under the Associationist Approach

Our three ''medicine'' figures illustrate a basic problem with the associationist approach when applied to causal situations: it does not allow a coherent definition of independent causal influence. Consider figures 1.3 and

1.4 again, but now treat medicines G and J as the same medicine. Under the ΔP model, the outcome patterns in the two figures are consistent with the medicine having no interaction with the background causes: the strength of the medicine according to the model remains invariant (-0.5 for both figures) across different levels of the background causes (as indicated by different proportions of patients having a headache in the control group).

Concluding independent causal influence based on the invariant ΔP value in the two figures, however, results in a contradiction in the concept of independent causal influence: the probability with which a cause changes an effect *varies* depending on the how often other causes of the effect occur, even as the causes are supposedly independent causes of the effect. As mentioned, the two figures show different contexts in which the medicine acted. In figure 1.3, as can be inferred from the control group (top panel) under the ''no confounding'' condition, generative causes of headaches occurred and produced headaches in every patient. In figure 1.4, it can be analogously inferred that generative causes of headache occurred and produced headaches in only half of the patients. Now, consider the probability with which the medicine prevented or relieved headache for an individual patient in the experimental group (bottom panel). Note that this group would have been like its control counterpart before the medicine was administered. In figure 1.3, a medicine that relieves headache with a probability of 0.5 for an individual patient would be expected to yield the bottom panel of the figure. But, in figure 1.4, to yield its bottom panel, the analogous probability for the medicine would have to be 1; individual patients are the units to consider because the pill that went into each patient cannot ''look'' across patients to obtain a desired overall outcome. Thus, the *same* medicine that relieved headache with a probability of 0.5 in figure 1.3 now relieved headache with a probability of 1, contradicting the ΔP assumption that the medicine acts independently of the background causes. It is as if the medicine, to maintain the same ΔP value, ''knew'' in figure 1.4 that it should bypass the patients who did not have a headache, and ''concentrate'' its influence on the half who did so as to arrive at a constant ΔP value of -0.5.

In summary, our analysis reveals that no model that treats figures 1.3 and 1.4 as symmetric patterns is coherent, given the asymmetry between the two values of a candidate cause. Adopting associationist models amounts

to assuming a specific form of interaction between the medicine and the background causes—the stronger the background generative causes, the weaker the preventive effect of the medicine—*even while* the medicine and the background causes are assumed to exert independent influences on headache. Such models violate the principle of invariance with respect to the basic concept of independent influence (Woodward 2003). The causal approach circumvents this problem by adopting the only coherent definition of independent causal influence.

Phenomenon 2: Causal-Reasoning Analogues of the Necker Cube

This section concerns an asymmetry between generative and preventive causes in situations in which $\Delta P = 0$ and there is no confounding. Recall that for such situations, both generative and preventive power may be evaluated. The general rule is to infer that the candidate is noncausal, regardless of whether generative or preventive power is evaluated. There are two exceptions, however, both of which are causal-reasoning analogues of the Necker Cube. Just as the Necker Cube, which results in a single visual input, can be interpreted in two ways perceptually, the exact same data set can yield two inferential interpretations. In the first analogue, if *e* always occurs, with or without *i* (see figure 1.5a), someone evaluating whether *i* prevents *e* would indeed infer that *i* is noncausal, but someone evaluating whether *i* produces *e* would allow the possibility that *i* produces *e*. Under the latter interpretation, even if *i* is a strong producer of *e*, there would be no room to show its influence on *e* because of the constant presence of *e* due to alternative causes; no conclusion, therefore, regarding whether *i* produces *e* can be drawn. This situation is referred to as the "ceiling effect" in experimental design.

A second analogue of the Necker Cube appears at the other extreme probability, when *e* never occurs, with or without *i* (see figure 1.5b). In that case, one would infer that *i* does not produce *e*; at the same time, *i*'s preventive power cannot be evaluated—there would be no room to show the preventive influence of *i* on *e*, no matter how strongly *i* prevents *e*. This situation is the preventive analogue of the ceiling effect; it is so exceedingly intuitive that it is, as far as we know, never discussed in textbooks on experimental design. Thus, each of two clear-cut cases of statistical independence is interpreted differently depending on the direction of

(a) Ceiling Effect

	e	not-*e*
i	1	0
not-*i*	1	0

(b) Preventive Analogue of the Ceiling Effect

	e	not-*e*
i	0	1
not-*i*	0	1

Figure 1.5
Two inferential Necker cubes: Data patterns indicating the ceiling effect and its preventive analog. Each cell entry is the P(column value | row value).

causality to be evaluated, and the interpretations for the two causal directions are reversed from one case to the other (for experimental evidence of this pattern of inference in college students, see Wu and Cheng 1999).

The differences in interpretation and the reversal in interpretation across the two cases are easy to overlook as crucial tests between associationist and causal accounts: being psychologically compelling rather than counterintuitive, the phenomena may suggest to some researchers that there is nothing interesting to explain. In fact, these phenomena pose unmet challenges to all associationist models, psychological or normative, and to some causal models as well; the compelling rationality of the pattern of judgments should render the gap in these accounts all the more conspicuous.

Associationist measures, such as the ΔP model and the chi-square test, always yield a single value for a data set; they therefore cannot yield two values for a ceiling situation. Some Bayesian network models make inferences regarding causal structure based on qualitative patterns of statistical independence and dependence (e.g., Spirtes, Glymour, and Scheines 1993/2000; Pearl 2000); these models would therefore also make the same inference, or set of inferences, for the same independence pattern. Some Bayesian network models and a version of the chi-square test exclude from

analysis data sets with extreme probability values; these models would be unable to explain why (under some conditions) it is actually fine to infer at $P(e) = 1$ that $p_i = 0$ and at $P(e) = 0$ that $q_i = 0$.

In contrast, these inferential Necker cubes are readily explained by equations 2 and 3 from the causal theory. If e occurs all the time, with or without i, q_i cannot be assessed (i.e., has the undefined value of 0/0 according to equation 2), as there are no remaining entities in which i can possibly manifest its generative power. But in this case, p_i would have a value of 0 according to equation 3—i never prevents e on the occasions when there is an opportunity to do so. Conversely, if e never occurs, p_i cannot be assessed (has the undefined value of 0/0 according to equation 3), because there are no entities in which i can possibly manifest its preventive power. But in that case, q_i would have a value of 0 according to equation 2.

Phenomenon 3: An Asymmetry between the ''Present'' and ''Absent'' Values of a Candidate Cause in Conjunctive Causal Inference

Implications for Inferring a Simple Cause versus Inferring a Conjunctive Cause

Recall that under our causal approach, a cause exerts its influence when it is present and does nothing when it is absent. At first glance, intuition may seem to blatantly contradict this assumption. There is no doubt that under some conditions people do speak of, and probably think of, the absence of factors as causes. For example, someone who is trying to quit smoking cigarettes might say, ''The absence of nicotine is causing my withdrawal symptoms;'' and someone who is separated from a loved one might say, ''Absence makes the heart grow fonder.'' But presence and absence are complementary concepts—what is expressible in terms of one can be similarly expressed in terms of the other. For example, ''The absence of nicotine causes withdrawal symptoms,'' is formally equivalent to, ''The presence of nicotine prevents withdrawal symptoms,'' although the two expressions may have different connotations. The choice of representation in the case of a simple cause does not affect whether one infers a causal relation.

In contrast, for conjunctive causes, whether someone forms the category of a conjunctive cause (i.e., whether one judges that the component causes interact) does depend on which value (presence or absence) of the component simple candidates is believed to be causal (Novick and Cheng 2004).

People's intuitions regarding conjunctive causes differ from the predictions of normative associationist models such as the chi-square test or the cross-product ratio, which would not yield different output values if the two values of each candidate factor are transposed (Novick and Cheng 2004). Neither would our associationist extension of the ΔP model to describe conjunctive causation (Cheng and Novick 1992). For both causal and associationist accounts, the respective symmetry properties of simple causes carry over to conjunctive causes because the latter are estimated on the basis of deviation from the independent influences of simple causes.

Let us illustrate the dependence of conjunctive causation on the causal value of a component factor. We will consider judgments on causal interaction in four patterns of outcomes, treating presence as a cause and absence as a cause in turn. For the next three figures, assume the following cover story:

> Scientists working for a company that raises a particular type of lizard for sale as pets are investigating factors that may influence the skin color of these lizards. In their natural habitats, lizards of this type have been found to have skin that is either yellow or black. These scientists have conducted an experiment to test the influence of two minerals, mineral *i* and mineral *j*, on the lizards' skin color. For each of four months, twelve lizards were exposed to mineral *i*, mineral *j*, neither mineral, or both minerals. At the beginning of each month, the lizards were given time to recover to their natural color before the relevant experimental manipulation began. All other influences on skin color were held constant throughout. The states of the lizards at the end of each month are depicted in the figure (with yellow represented by the lighter shade and black represented by black). Assume that the results are accurate and reliable, and suspend whatever prior knowledge you may have about lizards' skin colors.

Consider outcome pattern 1 (see figure 1.6). The four panels in this figure show the colors of the twelve lizards when the lizards were—respectively, from the top—exposed to neither mineral, exposed to mineral *i* alone, exposed to mineral *j* alone, and exposed to both minerals. A dominant interpretation of this pattern of outcomes is that the presence of the minerals independently causes some lizards to turn black (Liljeholm and Cheng 2005). Mineral *i* turned two lizards black; mineral *j* turned three other lizards black; and when exposed to both minerals, a lizard that was turned black by either mineral stayed black. This interpretation follows the principle of superimposition that is consistent with Novick and Cheng's (2004) theory of conjunctive causes: when two unlike values of an outcome are

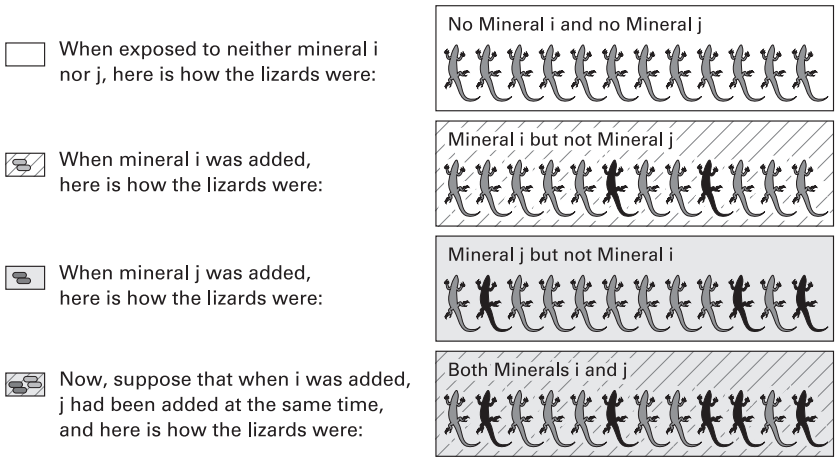

Figure 1.6
Outcome Pattern 1: An illustration of outcomes that show the independent influence of minerals I and J on the color of the lizards.

superimposed in an entity (e.g., when the varied outcomes in the second and sixth lizards from the top three panels are combined in the bottom panel), the value that wins (blackness) is the one that the entity newly adopted when a cause was introduced (i.e., when a mineral changed from being absent to being present).[2]

Pattern 2 (see figure 1.7) is identical to pattern 1 except that all twelve lizards are black in the bottom panel. For this pattern, the minerals each individually cause blackness as before, but they also interact to cause more blackness than would be expected if the minerals had operated independently.

Now, let us consider patterns that are ''absence'' analogues of patterns 1 and 2. Imagine two patterns, patterns 3 and 4, that are respectively identical to patterns 1 and 2 except that the exposure-condition labels (but not the lizards) for the top and bottom panels are reversed, so that the progression through the panels by convention specifies the removal rather than the addition of the individual minerals. Because the reversal focuses attention on the removal of the minerals, it should encourage the representation of the absence of a mineral as causal. It should be clear that pattern 3, being an absence analogue of pattern 1, should convey no causal interaction if absence is the causal value. This pattern should still indicate no interaction,

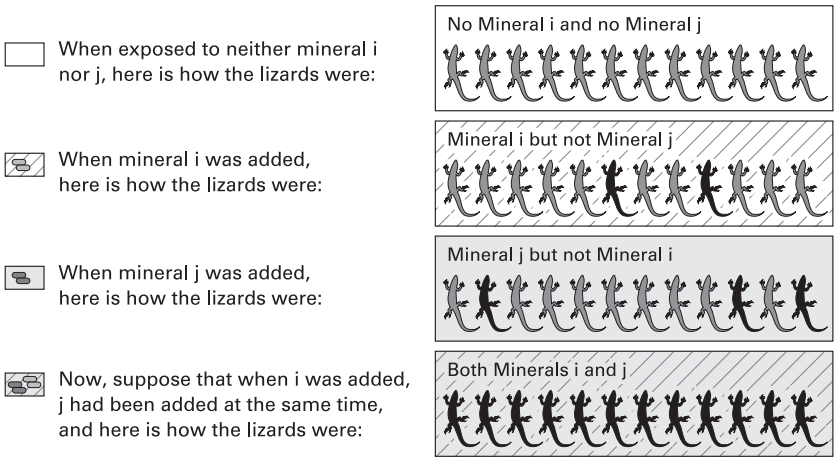

Figure 1.7
Outcome Pattern 2: An illustration of outcomes that show an interaction between Minerals I and J on the color of the lizards.

however, if presence is the causal value. In that case, mineral *j* turned two lizards yellow, mineral *i* turned three other lizards yellow, and when both minerals were present, their influences superimposed, turning five lizards yellow. Figure 1.8 shows pattern 4, the absence analogue of figure 1.7 (the background shadings that redundantly represent the exposure conditions are correspondingly shifted). If the *absence* of a mineral is perceived to cause blackness, pattern 4 should show an interaction of the absence of the two minerals to cause blackness, as does pattern 2 in terms of the *presence* of these minerals. Do reasoners spontaneously infer that the absence of the minerals interact in this case? In a similar vein, do reasoners infer that drugs that one is not taking interact with each other? We think not.

Our lizard figures were constructed to show that if only the ''present'' value of a candidate factor can be causal, pattern 4 would convey no causal interaction. Under this representation, the minerals each cause ''yellowness.'' Considering the panels in figure 1.8 from the bottom up, the presence of mineral *j* turned nine lizards ''yellow,'' the presence of mineral *i* turned ten lizards ''yellow.'' When both minerals were present, every lizard that was turned yellow by one or the other mineral, remained yellow (thus, all twelve lizards in the top panel were yellow). In other words, following

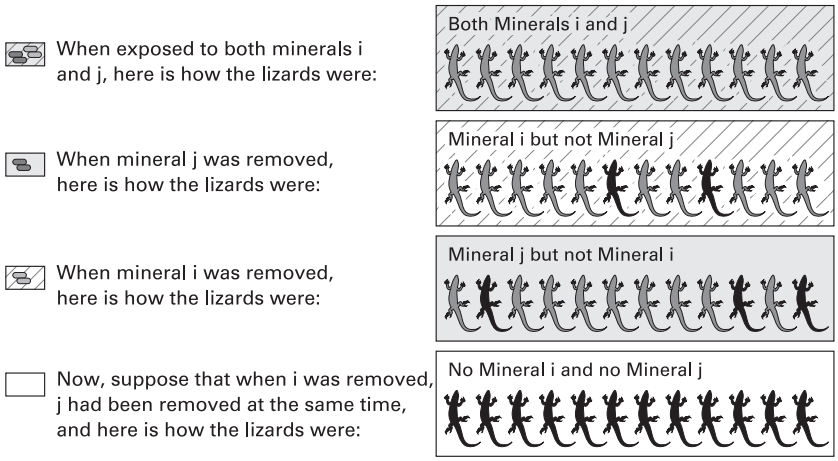

Figure 1.8
Outcome Pattern 4, the "absence" analogue of Pattern 2.

the same principle of superimposition as before—the value that wins is one that an entity newly adopted when a cause was introduced—the lighter shade now wins, and the pattern of superimposition (from the bottom up) conveys independence. Thus, the present and absent values of a causal factor are asymmetric, and judgments on causal interaction depend on this asymmetry.[3] Recall that, in contrast, predictions regarding causal interaction in associationist models are not dependent on the labeling of the two values of a candidate factor (Novick and Cheng 2004).

Empirical Evidence

To test whether only one value is causal and to measure untutored reasoners' default assumptions on the causal value, Liljeholm and Cheng (2005) conducted an experiment on college students. One practical difficulty to be overcome is that one cannot directly ask subjects about causal interactions without first essentially defining the concept for them, and potentially biasing their responses. To solve this problem, Liljeholm and Cheng showed subjects the four patterns and asked them to rate the complexity of the influences of the minerals; an interaction is more complex than independent influence. Because the ordering of the exposure conditions may contribute to complexity (e.g., the backwards ordering may feel

more complex), the critical comparison is between two differences in subjects' complexity ratings:

1. that between patterns 1 and 2 (the latter should be more complex if presence is considered causal); and

2. that between patterns 3 and 4 (the latter should be more complex if absence is considered causal).

The two differences should be comparable if presence is not marked as causal.

College student volunteers were randomly assigned to receive one of the four outcome patterns along with the cover story. Each subject was asked two related questions about the stimuli: ''What explains the changes in the lizards' skin color across the four panels?''; and, ''Given your understanding of the minerals' influences on the lizards, how complex do you think are those influences?'' Complexity, the dependent measure of actual concern, was specified as follows: ''If one panel—either the top or the bottom one in particular—is different from what you would expect given the influences of the minerals in the other three panels, the pattern would be complex.''

The subjects rated complexity on a numeric scale. If reasoners represent the presence of a factor as the causal value, the critical prediction is that the difference in complexity between patterns 3 and 4 would be smaller than that between patterns 1 and 2. A pattern of complexity ratings consistent with this prediction was observed: pattern 2 was (on average) rated considerably more complex than pattern 1 (more than 30 points higher on the 100-point scale); but patterns 3 and 4 were (on average) rated about equally complex (within 3 points of each other). The difference between the two differences was highly reliable. Recall that associationist models such as the ΔP and the chi-square predict no difference between the two differences.

In summary, at least in some situations, such as those in which the prevalence of the causal factors are unknown, the present value of a candidate factor is the one that exerts an influence despite the focus on the removal of the factor. The absent value does nothing, and the asymmetry between the two values is critical to whether two component factors form a conjunctive cause.

Phenomenon 4: An Asymmetry Around $P(e) = 0.5$ in Judgments of Whether a Causal Relation Exists

Overview

Another asymmetry occurs for answers to the basic qualitative question of whether a causal relation exists. When people are presented with data sets that have the same positive ΔP and are asked to judge whether a candidate cause produces an effect, and to rate their confidence in their judgment, their responses differ for data sets that are equidistant in opposite directions (i.e., symmetrical) around a 50/50 chance of the effect occurring. Consider two such data sets for which ΔP is 0.33: for one data set, on eighteen occasions on which *c* occurs, *e* occurs fifteen times; and on eighteen occasions on which *c* does not occur, *e* occurs nine times. For the other data set, on eighteen occasions on which *c* occurs, *e* occurs nine times; and on eighteen occasions on which *c* does not occur, *e* occurs three times. People are more confident that *c* causes *e* in the first data set than in the second (Liljeholm, Cheng, and Ford 2005). The chi-square test, in contrast, gives the same χ^2 value of 4.5 for both data sets, with the associated *p*-value of 0.03.

This asymmetry is interesting because it brings into focus the relations among structure learning, parameter estimation, and causal assumptions. Tenenbaum and Griffiths and others have drawn a distinction between structure learning and parameter estimation, and have characterized psychological work previous to their ''causal support'' model as concerning parameter estimation (Danks 2003; Griffiths and Tenenbaum, in press; Tenenbaum and Griffiths 2001). They write,

> The ΔP and [causal] power models correspond to maximum likelihood parameter estimates on a fixed graph (Graph_1), while the support model corresponds to a (Bayesian) inference about which graph is the true causal structure. (Tenenbaum and Griffiths 2001, caption for figure 1)

(See figure 1.9, below, for the graphs under consideration.) Griffiths and Tenenbaum (in press) further explain,

> Structure learning refers to identification of the topology of the causal graph, while parameter estimation involves determining the parameters of the functional relationships between causes and effects for a given causal structure. Structure learning is arguably more fundamental than parameter estimation, since the parameters can only be estimated once the structure is known.

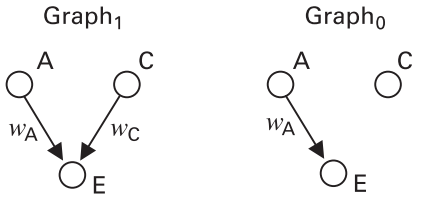

Figure 1.9
Two candidate causal structures representing, respectively, that C is, and is not, a cause of E.

In focusing on the topics of structure learning and parameter estimation, Tenenbaum and Griffiths have overlooked the causal assumptions that are at the heart of the debate between the associationist and causal approaches in the work on parameter estimation.

In fact, the causal assumptions that are critical to parameter estimation are just as critical to structure learning. To make this argument, it is necessary to begin with a description of Tenenbaum and Griffiths' (2001; Griffiths and Tenenbaum, in press) causal support model. (We limit our argument here to Bayesian inference, but a similar argument regarding causal asymmetry applies to null hypothesis testing.) As we will show, a causal variant of their model, which is consistent with the causal power theory, correctly predicts the asymmetry just described; in contrast, the other variant, an associationist one that is consistent with the ΔP model, incorrectly predicts symmetry. Ignoring the role of causal assumptions, Tenenbaum and Griffiths allow both variants in their model.

The Causal Support Model

The causal support model addresses the question of whether a causal relation exists (Griffiths and Tenenbaum, in press; Tenenbaum and Griffiths 2001); specifically, it evaluates which of the two causal structures in figure 1.9, Graph_1 and Graph_0, receives more support from the data. In the figure, C represents the candidate cause, and A represents alternative causes in the background. The parameters w_C and w_A are respectively the causal strength of C and of A to produce effect E. Graph_1 receiving greater support means that the evidence (the data) favors the existence of a causal relation between C and E. This model involves a comparison of the posterior probabilities of the two structures given the data—specifically, it takes the log of the

ratio of those probabilities—assuming that the structures have an equal prior probability and that all values of causal strength are equally probable prior to the consideration of the data, both for C and for A. Thus, in this model, the assessment of whether a relation is causal involves assumptions and computations regarding causal strength. The relevant data consist of information on event frequencies such as that illustrated in our earlier figures on the side-effects of medicines.

Tenenbaum and Griffiths (2001) show that the ΔP and causal power measures are each a maximum likelihood estimate of w_C in Graph_1 (see figure 1.9). These measures are consistent with two alternative functional relationships in the calculation of the likelihoods of the data given Graph_1 in the causal support model. The ΔP measure corresponds to a linear relationship between the probability of e in a situation and the strengths of the causes, with ΔP estimating w_C; the probability of e in the control group estimating w_A; and the probability of e in the experimental group estimating $w_A + w_C$. In contrast, causal power corresponds to a noisy-OR relationship (as explained earlier), with w_C corresponding to generative power, q_C; and w_A corresponding to $P(A) \cdot q_A$, the probability of e produced by alternative causes. In this case, the probability of e in the experimental group estimates $w_C + w_A - w_C \cdot w_A$. There is nothing in the Bayesian approach adopted by the causal support model that inherently restricts it to either functional relationship, provided that for the linear function values of $w_A + w_C$ outside the $[0, 1]$ interval are omitted.

The Role of Causal Assumptions

Tenenbaum and Griffiths' (2001; Griffiths and Tenenbaum, in press) characterization of the psychological debate as confined to different estimates of parameters under a fixed graph, and their suggestion that issues concerning parameter estimation are less fundamental than those concerning structure learning, are misleading. It might be tempting to assume that structure learning models such as Tenenbaum and Griffiths', making use of graphs such as those in figure 1.9 with their arrows so intuitively interpreted to depict causal relations, necessarily provide a causal analysis. A more accurate characterization of the debate and of the relations among the various proposed models (e.g., ΔP, causal power, and causal support), however, is that both structure learning and parameter estimation models are critically dependent on whether causal assumptions (such as those

underlying simple causal power) or associationist assumptions (such as those underlying ΔP) are made.

Even though the specific topic has been parameter estimation, the larger psychological debate has centered on whether causal assumptions are required, and the issues apply equally to structure learning. Like parameter estimation, structure learning can be causal or associationist. This distinction between causal versus associationist assumptions is critical because adopting the associationist linearity assumption implies that the structure learning model inherits the incoherent concept of independent causal influences discussed in the section on phenomenon 1.

The Asymmetry Revisited: Causal versus Associationist Predictions

As mentioned earlier, the assessment of whether a relation is causal in Tenenbaum and Griffiths' model (2001; Griffiths and Tenenbaum, in press), involves computations regarding the causal strengths of the candidate and background causes. When their model adopts the noisy-OR function (and corresponding causal assumptions) to combine causal strengths, it is able to explain the asymmetry in observed confidence judgments regarding whether a relation is causal mentioned at the beginning of this section; when their model adopts the linear function (and hence no causal assumptions), it is incapable of explaining these asymmetries. Let us return to the data sets that are symmetrical around the probability of 0.5, for which the chi-square test yields the same value of association. When the causal support model adopts the linear function, with values of $w_A + w_C$ restricted to the $[0, 1]$ interval, the model gives the same support value, namely 1.85, for both sets. In contrast, when this model adopts the noisy-OR function, it gives support values of 1.89 and 1.38 respectively, in qualitative agreement with observed human causal judgments (Liljeholm, Cheng, and Ford 2005).

Summary

The observed asymmetry around the probability of 0.5 of causal judgments on whether a causal relation exists, is consistent with the asymmetries in judgments of causal strength discussed earlier; the various asymmetries all stem from the causal assumptions underlying the judgments. An associationist process, by definition, does not make causal distinctions—such as the distinction between cause and effect, or the two values of a binary candidate cause—and thus gives rise to the various symmetric predictions;

predictions that reflect contradictory assumptions regarding independent causal influence. Thus, contrary to the view that assumptions about causal strength become relevant only after causal structure has been determined, these assumptions are in fact critical to a Bayesian evaluation of causal structure. Without the causal assumptions that underlie both parameter estimation and structure learning, neither the use of graphs nor Bayesian inference could constrain structure learning to be causal, or coherent.

The Appeal of Causal Representation: Coherence

The causal power theory (Cheng 1997, 2000; Novick and Cheng 2004) shows that there is coherence underlying people's intuitive causal judgments across multiple tasks. Current associationist models fail to capture this coherence in three related ways:

1. They do not allow for the possibility of independent causal influence.
2. They do not provide a parsimonious explanation of the four asymmetries; even with post-hoc settings of their current parameters, these models are still unable to account for the asymmetries.
3. They do not support the coherent derivation of other causal measures; in that sense, they fail to be *compositional*.

In contrast, the causal power theory (Cheng 1997, 2000; Novick and Cheng 2004) provides a parameter-free explanation of all four asymmetries. This unified explanation is a manifestation of the logical consistency and compositionality made possible by the causal assumptions under this approach.

A strong appeal of representing causal powers is that it allows the derivation of a variety of causal measures to answer different causal questions. We illustrated this capability in passing in our earlier sections. For example: (1) conjunctive causal power (Novick and Cheng 2004) is defined with respect to simple causal powers, making use of deviation from simple-power predictions; (2) generative simple power (equation 2) is logically related to preventive simple power (equation 3); and (3) the coherent variant of causal support (Tenenbaum and Griffiths 2001) incorporates the assumptions underlying simple causal power. Similarly, this approach allows one to derive answers to causal *attribution* questions; for example, what is the probability, given that an event has occurred, that the event is due to a particular candidate cause? Variants of attribution measures have appeared in the

literatures in psychology, law, epidemiology, philosophy, and artificial intelligence (e.g., Cheng and Novick 2005; Pearl 2000; Salmon 1965; Stott, Stone, and Allen 2004). The causal-power framework allows one not only to see how and why causal strength and attribution measures are related; but also to understand the conditions under which the measures hold, for example, why the estimation of all of these measures requires no-confounding. (For derivations of several attribution measures using causal power, including the probability of the necessity of a candidate cause to produce an effect, see Cheng and Novick 2005.) No-confounding, after all, is not always a requirement for causal inference (e.g., Haavelmo 1943). In this section, we give an intuitive illustration of these aspects of coherence with three measures—causal power; a causal attribution measure; and an interpretation of ΔP not as causal strength, as we have argued against, but as the probability with which an effect is produced by a candidate cause *i* alone when *i* is present.

We mentioned earlier the role of no-confounding in the derivation of equation 2, the estimate of generative causal power (Cheng 1997). For an intuitive translation of this role, consider the data illustrated in figure 1.10, from which one might wish to estimate the causal power of medicine B. Recall that causal power is the probability that a candidate cause *i* produces an effect *e* when *i* is present. If one could wear lenses that allow one to see the causes of an outcome, one would ''see,'' out of the patients in the experimental group (when medicine B is present), how many had headache caused by the medicine and arrive at the probability in question. Given the unavailability of causation lenses, some convoluted detective work is required: one first estimates the proportion of patients in the experimental group (the bottom panel) who would not have had a headache if they had not taken the medicine; then one observes the proportion out of this subgroup of patients who indeed have a headache, yielding the desired estimate. These patients' headaches must have come about as a result of medicine B and no other causes. One can estimate the first proportion by making use of the proportion of patients in the control group (the top panel) who do not have a headache (this is the denominator in equation 2; in the figure, this proportion is 2/3). But it should be clear from the nature of this estimate—in particular, from the use of the control group to make an inference regarding a quantity in the experimental group—that the estimate would be valid only if causes of headache other than the

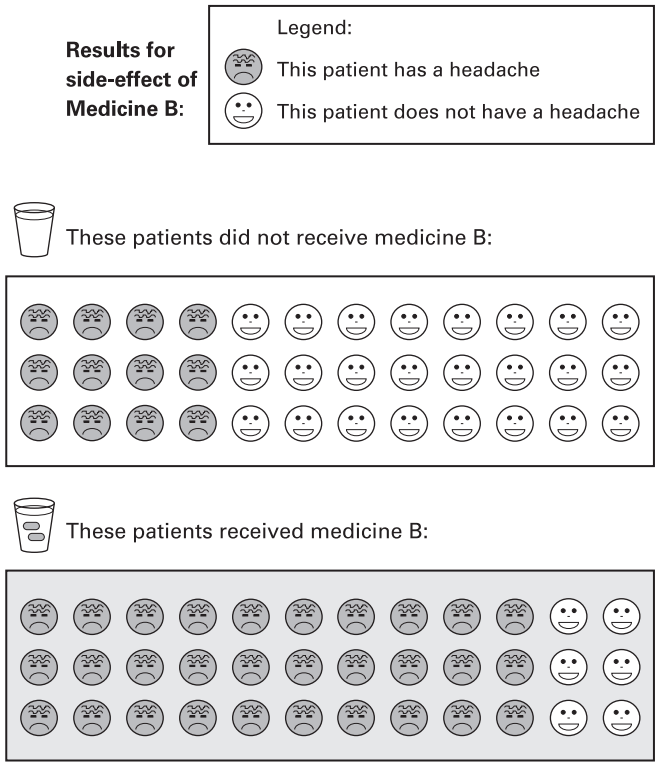

Figure 1.10
Stimulus material for Condition B from Buehner, Cheng, and Clifford (2003, Experiment 2).

medicine occur equally frequently in the two groups. This explains the no-confounding requirement for causal inference involving probabilistic contrast. In the figure, the second proportion—the proportion in the bottom panel who have a headache, out of those who would not have had a headache without the medicine—is 3/4. This is the generative causal power of medicine B.

Now, consider the probability that an effect e can be attributed to a candidate cause i, given that e has occurred, but not knowing whether i has occurred. (For psychological evidence for the use of this measure, see Johnson, Boyd, and Magnani 1994; and White 2004. For a more detailed explanation of the derivation of this measure, see Cheng and Novick 2005.) Let us denote this measure by $P(i \rightarrow e \,|\, e)$, where "$i \rightarrow e$" denotes that e is

produced by i. It is easy to see how causal power can serve as a building block for constructing this measure. Because e is caused by i with probability $P(i) \cdot q_i$, the measure is simply

$$P(i \rightarrow e \,|\, e) = \frac{P(i) \cdot q_i}{P(e)}, \tag{4}$$

with q_i as given by equation 2. Thus, the no-confounding condition required for estimating q_i is also required for $P(i \rightarrow e \,|\, e)$.

Although ΔP is incoherent as a measure of causal strength, it can be interpreted instead as $P(i\text{-alone} \rightarrow e \,|\, i)$, the probability with which an effect e is due to a particular cause i alone, knowing that i is present (but other causes of e may be present as well). This interpretation is different from causal strength in that it restricts attribution of e to i alone. The value of the estimate is therefore dependent on how often alternative causes of e occur in the context in question. Pearl (2000) refers to this as the probability that a cause is both necessary and sufficient to produce an effect. Our more explicit expression of the multiple relevant causal relations clarifies the apparent contradiction that arises from speaking of necessity and sufficiency in probabilistic terms.

As an estimate of $P(i\text{-alone} \rightarrow e \,|\, i)$, the ΔP expression can be derived using causal power. Let us return to figure 1.1. In the left box (when i is present), e is produced by i (as represented by the dashed circle) with probability q_i. On some of these occasions, e is also produced by alternative causes (as represented by the overlap of the two circles). The crescent part of the dashed circle (the complete dashed circle minus the overlap area) therefore represents the set of events in which e is produced by i alone. Recall that when there is no confounding, the undashed circle in this box, representing $P(a \,|\, i) \cdot q_a$, is estimated by $P(e \,|\, \bar{i})$. Thus, the desired probability is:

$$P(i\text{-alone} \rightarrow e \,|\, i) = q_i - q_i \cdot P(e \,|\, \bar{i}) = q_i \cdot [1 - P(e \,|\, \bar{i})] = \Delta P. \tag{5}$$

The simplification in the last step makes use of equation 2. Thus, the no-confounding condition that is required for estimating q_i is also required for $P(i\text{-alone} \rightarrow e \,|\, i)$. It should be clear that, under this interpretation, the ΔP value of a given candidate cause should vary depending on how large the undashed circle (and hence the overlap) is; in other words, it is not an invariant property of the candidate, as strength ideally should be. Other

measures of attribution can be similarly derived using the causal powers of candidate and alternative causes as elemental building blocks (Cheng and Novick 2005).

Summary and Conclusion

The four compellingly intuitive psychological asymmetries, which may appear unrelated at first glance, are in fact coherent under an approach that explains observable events by unobservable causal powers. Moreover, the theoretical construct of causal power—the ideally invariant probability with which a cause produces or prevents an effect—provides the building blocks that support the derivation of a variety of causal measures in answer to different causal questions, including:

1. measures for estimating the strength of simple causes and of conjunctive causes,
2. measures for evaluating the existence of causal relations, and
3. measures for estimating various causal attributions.

The coherence underlying the asymmetries, as well as the compositionality of causal power, reveals the incoherence of both psychological and normative associationist accounts. A central problem with the associationist approach is that, unlike the causal power approach, it does not allow the possibility of independent causal influence. The comparable problems that beset common statistical procedures have practical implications. Statistical inference is a cornerstone of science—and science distinguishes itself from quackery by its rationality. If some widely used associationist statistical measures suffer from a similar problem as their psychological counterparts —namely, the lack of causal representation—it would seem that science could benefit from a radical reassessment of associationist statistics when applied to test causal hypotheses.

Acknowledgments

The preparation of this paper was supported by NIH Grant MH64810. We thank Deborah Clifford for the graphic design of all of our figures, other than figures 1.5 and 1.9; Eric Ford for help with setting up integrals in our

calculations; and Clark Glymour and Keith Holyoak for clarifying our arguments. Correspondence concerning this article may be addressed to: Patricia W. Cheng, Mimi Liljeholm, or Christina Ford, Department of Psychology, Franz Hall, University of California, Los Angeles, California 90095-1563; or to Laura R. Novick, Department of Psychology and Human Development, Peabody no. 512, 230 Appleton Place, Vanderbilt University, Nashville, Tennessee 37203-5721. Electronic mail may be sent to: cheng@lifesci.ucla.edu, laura.novick@vanderbilt.edu, mlil@ucla.edu, or christis@ucla.edu.

Notes

1. No-confounding is not a necessary condition when there is information on more than two variables. (See the literature on instrumental variables, e.g., Glymour 2001; Haavelmo 1943; Pearl 2000; Spirtes, Glymour, and Scheines 1993/2000).

2. Note that the causal relations in pattern 1 do not satisfy the assumptions in Novick and Cheng's (2004) theory of conjunctive causes: specifically, the influences of the two component causes, minerals *i* and *j*, are not independent of the background causes, but instead are mutually exclusive (if mineral *i* changes the color of a certain lizard, mineral *j* does not, and vice versa). The pattern was so constructed as to allow both it and its absence analogue, pattern 3, to indisputably convey the independent influences of the two minerals. To explain the intuitive judgments, an extension of Novick and Cheng's theory that allows for an interaction between each candidate cause and the background causes would be needed.

3. Although we have focused our discussion on presence as the causal value, the absent value being causal would simply be the other side of the same coin. As long as only one value of a binary candidate factor is causal, the designation of the causal value would make a difference to whether a given situation involves conjunctive causation.

References

Buehner, M., P. Cheng, and D. Clifford. 2003. ''From Covariation to Causation: A Test of the Assumption of Causal Power.'' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29: 1119–1140.

Cartwright, N. 1989. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

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

Cheng, P. 2000. ''Causality in the Mind: Estimating Contextual and Conjunctive Causal Power.'' In *Explanation and Cognition*, edited by F. Keil and R. Wilson. Cambridge, Mass.: MIT Press.

Cheng, P., and L. Novick. 1992. ''Covariation in Natural Causal Induction.'' *Psychological Review* 99: 365–382.

Cheng, P., and L. Novick. 2005. ''Constraints and Nonconstraints in Causal Learning: Reply to White (2005) and to Luhmann and Ahn (2005).'' *Psychological Review* 112: 694–707.

Cheng, P., J. Park, A. Yarlas, and K. Holyoak. 1996. ''A Causal-Power Theory of Focal Sets.'' In *The Psychology of Learning and Motivation*, vol. 34: *Causal Learning*, edited by D. Shanks, K. Holyoak, and D. Medin. New York: Academic Press.

Danks, D. 2003. ''Equilibria of the Rescorla-Wagner Model.'' *Journal of Mathematical Psychology* 47: 109–121.

Danks, D., T. Griffiths, and J. Tenenbaum. 2003. ''Dynamical Causal Learning.'' In *Advances in Neural Information Processing Systems 15*, edited by S. Becker, S. Thrun, and K. Obermayer. Cambridge, Mass.: MIT Press.

Feinberg, S. 1980. *The Analysis of Cross-Classified Categorical Data*, 2d ed. Cambridge, Mass.: MIT Press.

Glymour, C. 2001. *The Mind's Arrows: Bayes Nets and Graphical Models in Psychology*. Cambridge, Mass.: MIT Press.

Griffiths, T., and J. Tenenbaum. In press. ''Elemental Causal Induction.'' *Cognitive Psychology*.

Haavelmo, T. 1943. ''The Statistical Implications of a System of Simultaneous Equations.'' *Econometrica* 11: 1–12.

Haavelmo, T. 1944. ''The Probability Approach in Econometrics.'' *Econometrica* 12 (supplement): 1–118.

Jenkins, H., and W. Ward. 1965. ''Judgment of Contingency between Responses and Outcomes.'' *Psychological Monographs* 7: 1–17.

Johnson, J., K. Boyd, and P. Magnani. 1994. ''Causal Reasoning in the Attribution of Rare and Common Events.'' *Journal of Personality and Social Psychology* 66: 229–242.

Lien, Y., and P. Cheng. 2000. ''Distinguishing Genuine from Spurious Causes: A Coherence Hypothesis.'' *Cognitive Psychology* 40: 87–137.

Liljeholm, M., and P. Cheng. 2005. ''Do the Absences of Things Interact?'' Unpublished manuscript, University of California, Los Angeles.

Liljeholm, M., P. Cheng, and C. Ford. 2005. ''Structure Learning, Parameter Estimation, and Causal Assumptions.'' Unpublished manuscript, University of California, Los Angeles.

Lober, K., and D. Shanks. 2000. ''Is Causal Induction Based on Causal Power? Critique of Cheng (1997).'' *Psychological Review* 107: 195–212.

Novick, L., and P. Cheng. 2004. ''Assessing Interactive Causal Influence.'' *Psychological Review* 111: 455–485.

Pearce, J. 1987. ''A Model for Stimulus Generalization in Pavlovian Conditioning.'' *Psychological Review* 94: 61–73.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Rescorla, R., and A. Wagner. 1972. ''A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement.'' In *Classical Conditioning II: Current Theory and Research*, edited by A. Black and W. Prokasy. New York: Appleton-Century Crofts.

Salmon, W. 1965. ''The Status of Prior Probabilities in Statistical Explanation.'' *Philosophy of Science* 32: 137–146.

Spirtes, P., C. Glymour, and R. Scheines. 1993/2000. *Causation, Prediction and Search*, 2d ed. Cambridge, Mass.: MIT Press.

Stott, P., D. Stone, and M. Allen. 2004. ''Human Contribution to the European Heatwave of 2003.'' *Nature* 432: 610–614.

Tenenbaum, J., and T. Griffiths. 2001. ''Structure Learning in Human Causal Induction.'' In *Advances in Neural Information Processing Systems 13*, edited by T. Leen, T. Dietterich, and V. Tresp. Cambridge, Mass.: MIT Press.

Van Hamme, L., and E. Wasserman. 1994. ''Cue Competition in Causality Judgment: The Role of Nonrepresentation of Compound Stimulus Elements.'' *Learning and Motivation* 25: 127–151.

White, P. 2004. ''Judgment of Two Causal Candidates from Contingency Information: Effects of Relative Prevalence of the Two Causes.'' *Quarterly Journal of Experimental Psychology* 57A: 961–991.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wu, M., and Cheng, P. 1999. ''Why Causation Need Not Follow from Statistical Association: Boundary Conditions for the Evaluation of Generative and Preventive Causal Powers.'' *Psychological Science* 10: 92–97.