Cognitive Psychology **40,** 87–137 (2000)
doi:10.1006/cogp.1999.0724, available online at http://www.idealibrary.com on IDEAL®

Distinguishing Genuine from Spurious Causes: A Coherence Hypothesis

Yunnwen Lien

National Taiwan University, Taipei

and

Patricia W. Cheng

University of California, Los Angeles

Two opposing views have been proposed to explain how people distinguish genuine causes from spurious ones: the *power* view and the *covariational* view. This paper notes two phenomena that challenge both views. First, even when 1) there is no innate specific causal knowledge about a regularity (so that the power view does not apply) and 2) covariation cannot be computed while controlling for alternative causes (so that the covariation view should not apply), people are still able to systematically judge whether a regularity is causal. Second, when an alternative cause explains the effect, a spurious cause is judged to be spurious with greater confidence than otherwise (in both cases, no causal mechanism underlies the spurious cause). To fill the gap left by the traditional views, this paper proposes a new integration of these views. According to the *coherence* hypothesis, although a genuine cause and a spurious one may both covary with an effect in a way that does not imply causality at some level of abstraction, the categories to which these candidate causes belong covary with the effect differently at a more abstract level: one

The research reported in this article was supported by NSF Grant DBS 9121298. This article is based in part on a Ph.D. dissertation completed by the first author under the advisorship of the second author. Experiment 1 was in part presented at the 33rd Annual Meeting of the Psychonomic Society, St. Louis, 1992 and at the Fyssen conference on Causal Understandings in Cognition and Culture in Paris, 1993. The presentation at the Fyssen conference appears as a chapter in D. Sperber, Premack, D., & Premack, A.J. (Eds.), *Causal cognition: A multidisciplinary debate*, Oxford: Oxford Univ. Press, 1995. We thank Michael Anderson, Bruce Burns, Deborah Decarvalho-Clifford, Catherine Fritz, Clark Glymour, Giyoo Hatano, Keith Holyoak, Frank Keil, Marc Lange, Douglas Medin, Michael Morris, Barbara Spellman, and three anonymous reviewers for their extremely helpful comments on our project. We thank Aileen Lee, Paul Rosen, David Shpall, and Damien Sim for conducting the experiments. Requests for reprints may be sent to Yunnwen Lien at the Department of Psychology, National Taiwan University, Taipei, Taiwan, or to Patricia Cheng, Department of Psychology, Unversity of California, Los Angeles, CA 90095-1563. E-mail: ywlien@cc.ntu.edu.tw; cheng@lifesci.ucla.edu.

0010-0285/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

covariation implies causality; the other does not. Given this superordinate knowledge, the causal judgments of a reasoner who seeks to explain as much as possible with as few causal rules as possible will exhibit the properties that challenge the traditional views. Two experiments tested and supported the coherence hypothesis. Both experiments involved candidate causes that covary with an effect without implying causality at some level, manipulating whether covariation that implies causality has been acquired at a more abstract level. The experiments differed on whether an alternative cause explains the effect. © 2000 Academic Press

INTRODUCTION

It is often believed that atmospheric pressure drops before the approach of a storm. It is also often believed that ants and other underground insects migrate to higher places before a storm. Now, whereas people would typically accept that a drop in atmospheric pressure causes storms, they would not think that the migration of ants *causes* storms, even though they may make use of observed migration to predict storms. In a similar vein, if one is told that all members of some school board became bald soon after joining, whereas most other people did not, one is unlikely to believe that being a member of that board causes one to be bald. In each of these cases, one phenomenon is *associated* with a second: when the first occurs, the second is likely to occur. The first phenomenon also *covaries* with the second: when the first occurs, the second is more likely than otherwise. Of these regularities, some but not others are likely to be judged as causal. How do people judge whether a regularity is causal?

We call a factor that covaries with an effect (or is associated with it) and is judged to be causal a *genuine* cause and a factor that covaries with an effect (or is associated with it) but is judged to be noncausal a *spurious* cause (terms suggested by Suppes, 1970). Spurious causes may be classified into two categories. The regularity between some spurious causes and the effect can be plausibly explained by an alternative cause. For example, the storms (the effect) that follow the migration of ants (a spurious cause) is due to a drop in atmospheric pressure—an alternative cause that explains both the storms and the migration. Other spurious causes are not clearly explained by any alternative cause, but might covary with the effect purely by chance. The school board example is of that kind: the members might just have coincided in becoming bald. People are likely to be more confident of their judgment that a candidate is a spurious cause when they know of an alternative cause that explains the regularity than when they think the regularity is accidental. It is possible to imagine, for example, that members of the school board did not all become bald by coincidence, but that instead, membership on that board (the candidate cause) starts a causal chain that results in baldness: membership causes stress, which in turn causes premature aging, with baldness as a symptom. For the unexplained kind of spurious cause, without

additional information that rules out possible causal paths between the candidate and the effect, a reasoner cannot be confident that the regularity is accidental.

Both psychological research and our common sense tell us that people systematically distinguish genuine from spurious causes (e.g., Bullock, 1979; Bullock, Gelman, & Baillargeon, 1982; Leslie & Keeble, 1987). In the present article, as the preceding examples illustrate, we are concerned with the distinction *under the situation in which not all alternative causes are believed to occur independently of the candidate cause*, so that the covariation in question does not in itself imply causality (see Cheng, 1997, for a theory of why causal inference is possible under certain conditions if alternative causes *are* believed to occur independently of the candidate cause). In such situations, when two candidate causes covary equally with an effect, and one candidate is judged to be causal and the other not, other knowledge must explain the difference. Most researchers agree that this additional knowledge is causal in nature.

The Criterion of Causal Power

Power theorists have argued that the knowledge of some causal mechanism or the transmission of some causal power due to a generative source—knowledge that goes *beyond* covariation—is essential for people to distinguish genuine from spurious causes (e.g., Ahn, Kalish, Medin, & Gelman, 1995; Bullock et al., 1982; Harré & Madden, 1975; Koslowski, 1996; Salmon, 1977, 1984; Shultz, 1982; Shultz & Kesterbaum, 1985; White, 1989). According to this view, when an underlying causal power or mechanism is perceived or known, a covariation will be judged as causal; otherwise, it will be unlikely to be judged so.

The power view is well illustrated with an experiment by Bullock (1979; Bullock et al., 1982). In this experiment, preschool children judged what caused a Jack-in-the-box to pop up from its box. Before it popped, two identically timed events simultaneously occurred: a ball rolled down a slope toward the box and a series of lights ''rolled'' down a parallel slope toward the box. Both the ball and ''the light'' were occluded from sight briefly before ''Jack'' was seen to pop. The above sequence was repeated multiple times, thus defining identical covariational relations for the rolling of (a) the ball and (b) the light with respect to Jack's popping. Note that because each covariation is confounded by the other, neither implies causality.

In one condition, the entire apparatus—which consisted of three separate boxes respectively containing the ball, the lights, and Jack—appeared to be in a single box. Bullock et al. (1982, p. 225) reasoned that the popping of Jack was thus consistent with a plausible mechanism involving impact by the ball: ''rolling and hitting could produce movement in another object through impact.'' In contrast, there was no plausible causal mechanism underlying

the covariation involving the traveling lights. A consideration of mechanisms should therefore lead a subject to attribute the popping to the ball but not the light, despite identical covariations for the ball and the light.

In another condition, a 6-inch gap was seen between the box containing Jack and the boxes containing the ball and the light. The popping of Jack was thus inconsistent with the mechanism for producing movement through impact. Bullock et al. argued that older children who were familiar with electrical phenomena might be more likely to attribute the popping to the light. Such children might reason that the light was produced by electricity, which can travel through the gap by hidden wires and hence provide a plausible mechanism for Jack to pop. Their predictions were confirmed.

Bullock et al.'s (1982) experiment cleverly demonstrated a situation in which a covariation does not in itself imply causality, allowing them to make the important point that prior causal knowledge is critical to causal judgments in such situations. Because causal relations that are perceived to be truly novel (i.e., not instances of any known kinds of causal relations) are probably rare, understanding the influence of prior knowledge of mechanisms is important for understanding everyday causal inference (Ahn et al., 1995).

Incompleteness of the Power Explanation

The power explanation of the distinction between genuine and spurious causes is obviously incomplete, however. We note three problems (also see Cheng, 1993; Glymour & Cheng, 1998).

Where does knowledge of causality come from? First, the power explanation begs the question: Where does knowledge about the *causal* nature of mechanisms come from? For example, Bullock et al. (1982) simply assumed that the children knew the causal mechanism that subsumes the relation between the ball's rolling and Jack's popping—impact by a moving object. How knowledge regarding its causal nature came about was left mysterious. Because the causality of causal mechanisms is ultimately where their explanatory power lies, a more complete explanation would include an account of how knowledge of causal mechanisms comes about.

Level of abstraction of a causal mechanism. To explain the distinction between genuine and spurious causes by the perceived existence of a causal mechanism, one must specify the level of abstraction at which that mechanism is represented: Whether one perceives that a causal mechanism underlies a covariation depends critically on that level. (We use ''level of abstraction'' to refer to degree of inclusiveness or level of generality.) If a causal mechanism is represented more narrowly than is objectively true, the reasoner could fail to perceive that it underlies a covariation in question. For example, if the cause of mechanical motion were represented overly narrowly as ''impact by *solids* with momentum,'' the movement of a leaf in the wind (wind is not solid) would be misperceived as having no ''underlying causal mechanism.'' In other words, the wind would be misperceived as a

spurious cause of the movement of the leaf, as there would be no superordinate of wind that is a cause of the movement of objects. Likewise, if inhaling fumes from tobacco as a cause of lung cancer were represented overly narrowly as ''inhaling fumes from Virginia Slims,'' then a covariation between smoking Camel cigarettes and lung cancer would be perceived as spurious.

Conversely, if a causal mechanism is represented more generally than is objectively true, the reasoner could misperceive that mechanism to underlie a causally irrelevant covariation. If the cause of mechanical motion were represented overly generally as ''impact by an entity,'' then a contrived covariation, for example, between photons from a light shining on a leaf and the leaf's movement, would be misperceived as having an underlying causal relation and hence as causal. Likewise, if the inhalation of all fumes were overgenerally represented as a cause of lung cancer, then a contrived covariation between inhaling steam and lung cancer would be misperceived as causal.

As people do not have the above misperceptions, they cannot be representing the cause of mechanical motion as either impact by only solids with momentum or impact by an entity; nor can they be representing a cause of lung cancer as either the inhalation of fumes from a brand of cigarettes or inhalation of all fumes. What determines the level of abstraction at which a causal mechanism is represented so that such misperceptions would not result?

Difference in confidence in judging explained and unexplained spurious causes. Even given the appropriate level of abstraction of prior causal knowledge, the power view still cannot explain the difference in confidence in judging explained and unexplained spurious causes to be spurious: An underlying causal mechanism is *equally* absent from the spurious causal relation whether or not this relation is explained by an alternative cause.

Integrating and Extending the Power and Covariation Views

An answer to how causal knowledge comes about is given by Cheng (1997), whose theory extends what has been called the *covariation* view, an approach to the discovery of causal relations based on observable information alone (see Shanks, Holyoak, & Medin, 1996), by incorporating the concept of causal power. Although Cheng's theory explains a diverse range of findings that are inexplicable by accounts that exclusively take a power or a covariation approach (see Cheng, 1997, for a review), people are often able to make causal judgments in situations in which her theory does not apply, as we explain later. Our coherence theory, which makes use of the optimal representation of causal relations, aims to further integrate the power view with the covariation view to provide a complementary solution to the above questions. We introduce our theory preceded by Cheng's causal power theory and a proposal on the optimal representation of causal relations.

Origin of causal knowledge: explaining conditional contrasts by causal powers. Cheng (1997) argues that reasoners postulate the existence of general types of unobservable causal powers. The *generative power* of a candidate cause *c* with respect to effect *e* is the probability with which *c* produces *e* (Cartwright, 1989). The *preventive power* of *c* is the probability with which *c* prevents an otherwise-occurring *e* from occurring. To estimate specific unobservable powers, reasoners bootstrap by hypothetically using these powers to explain covariation, which is defined purely in terms of observable information. The covariation model explained by causal powers in Cheng's theory is the probabilistic contrast model (Cheng & Novick, 1990), which applies to potential cause and effect events that can be represented as binary variables, with the cause events perceived to precede the effect events.[1] According to this model, given that *c* is perceived to occur before *e*, reasoners assess the *probabilistic contrast* for *c* with respect to *e*, ΔP_c, over a *focal set*, where

$$\Delta P_c = P(e|c) - P(e|\bar{c}) \tag{1}$$

and ''focal set'' is the set of events that the reasoner uses as input to the covariation process. A mental construct in the reasoner's explanation is a distinction between *c* and the composite of (known and unknown) causes alternative to *c*, which we label *a*. For example, when reasoners evaluate whether *c* produces *e*, they explain $P(e|c)$ in Eq. (1) by the probability of the *union* of two events: (1) *e* produced by *c* and (2) *e* produced by *a* if *a* occurs in the presence of *c*. That is, they reason that when *c* is present, *e* can be produced by *c* or by *a* if *a* occurs in the presence of *c*. Likewise, reasoners explain $P(e|\bar{c})$ in that equation by how often *e* is produced by *a* alone when *a* occurs in the absence of *c*. These explanations yield equations that allow the estimation of the power of *c* in terms of observable frequencies under some boundary conditions.

A mathematical result in Cheng's (1997) theory is that to infer a new causal relation, one does not need to know what other causes of the effect are; one only needs to know that, whatever these causes may be, they occur independently of the candidate cause, for example, by being constant. We refer to contrasts in which alternative plausible causes are controlled as *conditional contrasts*. For nonnegative conditional contrasts ($\Delta P_c \geq 0$), Cheng's derivation shows that

$$p_c = \frac{\Delta P_c}{1 - P(e|\bar{c})}, \tag{2}$$

[1] Explaining other covariation models yields mathematically equivalent solutions for causal power.

where p_c is the generative power of c with respect to e. An analogous equation results for nonpositive conditional contrasts in the case of the evaluation of preventive power.

As mentioned, despite the strong support for conditional contrast as a criterion for discovering causes, in many situations in which causal judgments are made, this criterion cannot be applied. People may not have sufficient information for computing such a contrast. For example, a candidate cause may covary perfectly with a genuine cause, in which case information is not available for estimating one of the two probabilities in the conditional contrast for the candidate cause. Take the example of the covariation between the migration of ants to higher places (the candidate cause M), and the approach of a storm (the effect S). The migration of ants may covary perfectly with L, low atmospheric pressure (an alternative cause that needs to be held constant). Thus, while $P(S|LM)$—the probability of a storm approaching, given that atmospheric pressure is low and the ants migrate—is available (it is high), $P(S|L\overline{M})$—the probability of a storm given low atmospheric pressure and the *absence* of migration—is undefined: when the atmospheric pressure is low, the ants never fail to migrate. The analogous contrast for migration conditional on atmospheric pressure being high likewise cannot be computed. Bullock et al.'s (1982) experiment described earlier is another example of the same problem. More generally, for a variable at a particular level of generality, it is often not possible to find a set of events in which alternative causes occur independently of it.

Nonetheless, even when conditional contrast cannot be computed, people often seem capable of making a systematic distinction between genuine and spurious causes. For example, despite never having encountered ants migrating in the wrong weather, people are unlikely to judge that their migration is the cause of an approaching storm. If it is people's prior causal knowledge that allows them to make such a distinction, how does that knowledge give rise to the various genuine and spurious causal judgments? We return to answer this question after the following proposal.

A criterion for determining the level of abstraction. In his discussion of chaos, Lewis (1929) noted that the very fact that people represent the world in terms of categories denies the possibility of complete chaos: *categories are what obey laws.* Our solution to the problem of the level of abstraction of a causal relation makes use of Lewis' idea. Under the obvious constraint that the levels under consideration are available to the reasoner, we suggest that the adopted level is the one at which the contrast for the candidate cause with respect to the effect in question is at its maximum (Cheng, 1993). The state of the cause then optimally predicts the state of the effect. To illustrate our point regarding the optimal level of representation of a cause, or cause category, let us consider the simple case in which c is a deterministic cause of effect e, and all alternative causes, represented by the composite a, are absent. The conditional contrast for c in this case is $P(e|c\bar{a}) - P(e|\overline{c}\overline{a})$. For

clarity of exposition, we drop the notation for $\bar{a}$ in the rest of this section and simply assume that context, noting that *a* remains the same set of causes throughout this explanation.

Ideally, the contrast value will equal 1 (the maximum). This occurs when *c* is defined at a certain level of abstraction so that instances of *c* invariably elicit *e* [i.e., $P(e|c) = 1$) and non-*c* instances never elicit *e* (i.e., $P(e|\bar{c}) = 0$). When *c* is defined more generally than that level, some instances that do not elicit *e* are included under this category. That is, some instances that are classified as $\bar{c}$ in the ideal case (and therefore do not elicit *e*) are now classified as *c*. Hence, the value of $P(e|c)$ is less than 1 [whereas $P(e|\bar{c})$ is still 0]. As a result, the contrast is less than 1. Conversely, when *c* is defined more specifically than the ideal level, some causal instances are excluded as such. That is, some instances that are classified as *c* in the ideal case (and therefore elicit *e*) are now classified as $\bar{c}$. The value $P(e|\bar{c})$ is therefore greater than 0 [whereas $P(e|c)$ will still equal 1], and the contrast will also be less than 1. Therefore, when *c* is not represented at the optimal level of abstraction, its contrast value will be less than its maximum.

To apply the criterion of maximal contrast, there must be a predefined effect at a certain level of abstraction with respect to which the contrast for a candidate cause can be computed. In this article we assume that there are effects at particular levels of abstraction (e.g., lung cancer, the movement of objects, and baldness) that are of a priori concern to the reasoner. Note, however, that the cause category, the one at the optimal level, is a *result* of this criterion. In contrast to previous accounts of causal induction, in which the definition of candidate causes is a prerequisite, this criterion implies that (1) what defines a candidate cause and (2) whether it causes an effect involve a *single* decision. The previous prerequisite is therefore unnecessary.

Let us now qualitatively illustrate the concept of maximal contrast for a nondeterministic causal relation. If ''smoking cigarettes'' is indeed a cause of lung cancer, then one would expect its conditional contrast to be reduced when the cause category is defined either more generally or more specifically. Consider defining the cause category more generally as ''inhaling fumes.'' Because lung cancer does not occur after inhaling steam and other harmless fumes, *P*(lung cancer|inhaling fumes) would be less than *P*(lung cancer|smoking cigarettes). Recall that the same set of alternative causes are held constant for contrasts at the various levels under comparison. The probability of lung cancer should therefore remain unchanged conditional on ''no cigarette smoking'' or ''no fume inhaling'' (the same alternative causes produce lung cancer in both cases). It follows that the contrast with respect to lung cancer is reduced when the cause category is defined more generally. Now, consider defining the cause category more specifically as ''smoking Virginia Slims.'' Assume for simplicity that all brands of cigarettes cause lung cancer just as much. Because lung cancer often occurs after smoking other brands of cigarettes, *P*(lung cancer|no Virginia Slims) would

be larger than *P*(lung cancer|no cigarette smoking). It follows that the contrast with respect to lung cancer is also reduced when the cause category is defined more specifically.

The same argument applies to the level of representation of other causal relations, for example, to explain why ''allergy to dairy products'' is the common representation for that type of allergy rather than the more specific representation of ''allergy to cheese'' only or the more general one of ''allergy to products from cattle,'' ''allergy to foods from animal sources,'' and so on. Likewise, if greater precision is desired, and the relevant morbidity information is available, the same argument can be made to explain more precise definitions of causes, for example, ''inhaling tar'' rather than ''smoking cigarettes'' as a cause of lung cancer and ''lactose'' rather than ''dairy products'' as the source of the allergy. If these more precise representations are not commonly used, it is not because they are perceived as incorrect or odd, as our examples involving reduced contrast would be, but because they involve entities that are unobservable with the tools available in everyday life.

Our maximal-contrast criterion is a computational-level description (Marr, 1982) of the level of abstraction at which a causal relation is represented. It specifies that level without any commitment to the algorithm by which reasoners arrive at it. We return to the issue of what constitutes an alternative cause in the General Discussion.

The Coherence Hypothesis. Like our maximal contrast criterion, our *coherence* hypothesis is a computational-level description. To make use of the power view, we first rephrase it. A novel event or object at a specific level of abstraction (e.g., ball) may have one or more features that renders it an instance of a known kind of event or object at a more abstract level (e.g., object with mass). A novel covariation regarding this event or object (e.g., under a given context, if and only if a ball hits another object, this object takes off) is then an instance of a more abstract relation, sometimes a relation that is known to be causal (e.g., under a given context, when an object with mass hits another object with mass, the impact *causes* the latter object to take off). According to our interpretation of the power view, then, a genuine cause and a spurious one are members of different categories: whereas a genuine cause has a feature that renders it a member of a familiar causal category, a spurious cause does not. That feature may be any of a variety of causal concepts: a cause in a direct causal relation, an interpolating link in a causal chain, an enabling condition, or a component in a complex cause consisting of multiple factors. Thus, the rolling ball in Bullock et al.'s (1982) Jack-in-the-box experiment is also a moving object with mass, making the ball a member of a familiar category that can cause movement of other objects. Consider another example of a causal mechanism: Dave's stomach problem after eating chicken at a local restaurant is explained by the restaurant's chef undercooking chicken (Ahn et al., 1995). In this example, one

of Dave's features—salmonella in his stomach—forms a link in a chain of events (the chef undercooking chicken, the salmonella on the chicken remaining alive, Dave eating chicken with salmonella at the restaurant, salmonella in Dave's stomach, Dave's illness). That feature is an instance of a familiar category that can cause stomach problems. Considering the consistency between a covariation in question and more abstract causal relations takes us one step closer to our goal: For abstract as well as specific binary variables, it is possible to infer the variety of causal concepts listed earlier from observable frequencies (Cheng, 1997; Spirtes, Glymour, & Schienes, 1993).

We now take the further step of explaining why consistency matters. Let us start with an existing knowledge system and consider whether there is pressure to add a causal rule to explain an observed novel covariation. Assume that (1) causal rules explain covariations and (2) a reasoner seeks to explain as much as possible with as few rules as possible. On one hand, if any of the categories to which a candidate cause belongs is a familiar cause of the effect, the covariation involving the candidate would be consistent with this superordinate causal relation. We term consistency of this sort *hierarchical consistency*. Judging such a candidate to be causal would not require any change to the existing knowledge system, because a hierarchically consistent covariation is subsumed, explained, or predicted by the more general causal relation. In contrast, judging such a candidate to be noncausal would require the creation of an exception rule at the same time that the covariation is left unexplained (i.e., the covariation would be merely incidental). The coherence hypothesis therefore predicts that a hierarchically consistent covariation is likely to be judged causal.

On the other hand, if the covariation in question is hierarchically inconsistent (i.e., no superordinate of the candidate is a cause of the effect), it would be unexplained or unpredicted by the candidate. In this case, if an alternative candidate that covaries with the effect in the given context is known to be a cause (e.g., a drop in atmospheric pressure in the storm example), the covariation in question would be consistent with preexisting rules regarding that alternative cause. The coherence hypothesis therefore predicts that the covariation between the candidate and the effect would be judged as spurious: judging it so requires no new rule and leaves no covariation unexplained, whereas judging it as causal requires adding a new causal rule without explaining more.

Otherwise (i.e., in the absence of a covarying alternative candidate that is a cause of the effect), the hierarchically inconsistent covariation is unexplained by the existing knowledge system. In this case, there is no happy solution: The unexplained covariation would create pressure for a new causal rule, for which there is no evidence—conditional contrast cannot be computed. Because the reasoner seeks to explain with as few rules as possible, no new causal rule is created (i.e., the covariation is deemed spurious), but

the unmet pressure for explanation undermines confidence in the reasoner's conclusion.

Illustrations. Let us illustrate the coherence hypothesis by returning to Bullock et al.'s (1982) experiment. Although information is unavailable for computing conditional contrast for either the rolling ball or the traveling light with respect to Jack's popping, the ball, in contrast to the light, is a moving object with mass. The children were likely to have independently manipulated impact (e.g., pushed an object with their body or hand or rolled a ball that proceeded to hit another object) or have seen other people manipulate it. Information should therefore be available for computing conditional contrast (e.g., when there is impact by a moving object with mass on another object, motion in the latter object is more likely than when there is no such impact, other things being equal), allowing a causal rule to be inferred regarding impact, such as the one stated earlier. When there was no gap between the rolling ball and Jack, the covariation between them is consistent with this general causal knowledge (there is impact) and is therefore likely to be judged as causal. In contrast, when there was a gap, that covariation is inconsistent with this knowledge (there is no impact) and is therefore likely to be judged as noncausal.

Let us also illustrate our hypothesis with an example of a spurious causal relation in real life, returning to our ants and storms example. Recall that there is insufficient information for computing the conditional contrast for migration with respect to storms. At a more abstract level, however, plenty of information on manipulating movements of objects does exist, allowing causal knowledge to be inferred at that level. For example, there is likely to be sufficient observable information for learning that the influence of a force depends on the location and direction in which it is applied and that for movements of the same speed, the small force created by the movement of light objects can move only light objects. The covariation between the movement of ants and storms is unlikely to be subsumed by any such knowledge: the movement of large objects (e.g., masses of air and rain in the sky during a storm) is caused by large forces applied in the right location and direction (e.g., the suction produced by low atmospheric pressure), a category to which the ants' migration on the ground does not belong.

Note that it is possible to have insufficient information for computing a specific conditional contrast (e.g., ants' migration and storms) and yet have sufficient information for *other* specific conditional contrasts, allowing specific causal inferences (e.g., vacuuming causes air and dirt to be sucked into the vacuuming cleaner, a narrow passage between buildings causes air to rush through the passage) and inference to the kind (gas rushes from areas where the pressure is high to fill those areas where pressure is low). This assumption is not specific to our hypothesis. In order for acquired general causal knowledge to have an influence on specific causal judgments, it must be assumed that it is possible to have sufficient information for acquiring

an abstract causal mechanism, even when there is insufficient information for judging the causality of a more specific relation at its own level per se.

Our framework provides a definition of the knowledge of an underlying mechanism: An acquired causal mechanism need be nothing more than a causal relation based on a conditional contrast that subsumes a covariation in question. This demystification both raises and answers the related critical questions of how and at what level of abstraction the requisite causal knowledge is acquired.

Scope. As implied by the preceding discussion, this article concerns causal relations involving candidate causes and effects that are represented by binary variables or by other types of variables that can be recoded into that form.

In many cases our causal knowledge is no doubt culturally transmitted. In such cases our question is simply pushed back in time: How did people first come to judge whether a covariation was causal? In some domains, an innate ability to distinguish genuine from spurious causes may exist (e.g., Garcia & Koelling, 1966; Leslie & Keeble, 1987). Much of causal knowledge, however, is probably acquired. It seems implausible, for example, that an innately known causal mechanism explains why people attribute storms to low atmospheric pressure rather than the migration of insects. This article focusses on acquired causal knowledge.

New causal relations are sometimes learned by analogy to a known causal relation. Because our interest here is in tracing the origin of causal knowledge, and the causality of an analog is traceable to the known relation, we do not treat causal learning by analogy as a separate case, even though this learning is likely to involve a nontrivial mechanism (Gentner, 1983; Hummel & Holyoak, 1997).

Rationale for Experimentation

We noted that although a genuine cause and a spurious one may both covary with an effect in a way that does not imply causality at some level of abstraction, these candidate causes have different causal status at a more abstract level: one is causal, the other is not. Among spurious causal relations, if reasoners seek to explain as much as possible with as few causal rules as possible, then they should judge a covariation to be noncausal with greater confidence when they know of an alternative cause that explains it than when they do not. This difference has not been explained or tested previously.

To test our explanation, we created a situation in which the covariations presented are not conditional on constant values of alternative causes, but superordinate causal knowledge is available—knowledge inferred from conditional contrasts superordinate to the covariations in question. *For candidate causes that do not occur independently of alternative causes of an effect, can consistency between the covariation involving the candidate and more*

abstract causal knowledge inferred on the basis of conditional contrasts explain the various kinds of judgments regarding genuine and spurious causes? We report two experiments in answer to this question: The first tests the case in which subjects did not know of an alternative cause of the effect, and the second tests the case in which they did.

EXPERIMENT I: UNEXPLAINED SPURIOUS CAUSAL RELATIONS

All subjects first read a cover story, after which they were given a learning task (this was the *learning* phase). In the cover story, they were asked to imagine that they were applying for a job as an assistant to a gardener. The gardener gave each applicant a test to see how well he or she could discover what caused a certain kind of flowering plant to bloom. He told them that each group of 10 plants of this type grown in his yard was fed with a different substance. In addition, 10 plants were not fed any substance. He put all the plants in his greenhouse to keep the environmental conditions constant. He then showed the applicants whether each of these plants subsequently bloomed. For example, most of the unfed plants did not bloom. This information allowed subjects to infer abstract causal relations based on conditional contrasts regarding the substances shown.

All subjects were then given the gardener's test (in the *test* phase). They were shown that in a novel environment (a possible alternative cause), when several plants of the same type were fed *s*, a novel type of substance, most of these plants bloomed. Subjects' main task was to judge whether *s* was a cause of blooming. In particular, we assessed whether subjects thought that most plants of this type—if fed *s* in the *old* environment—would bloom (i.e., whether $P(\text{blooming}|s)$ in that environment would be large). Our assessment concerned subjects' predictions regarding the outcome of an intervention involving the candidate cause. If subjects interpreted the association or covariation between *s* and blooming to imply that *s* causes blooming, they should predict that most such plants would bloom.

Note that although *s* covaried with blooming, it also covaried perfectly with the environment. As a result, there was insufficient information for computing the relevant conditional contrast for *s:* The frequency of blooming in the absence of *s* conditional on the novel environment was unavailable. The situation we created was one in which no known or highly plausible cause was available to explain the occurrence of the effect when the target association was inconsistent with prior causal knowledge. Although the possibility of the new environment as the cause could not be ruled out, no positive evidence was presented to support this possibility.

Our only manipulation was the hierarchical consistency between the inferred abstract causal relation and the covariation between *s* and blooming. During the learning phase, although every subject was shown an identical set of various substances fed to plants of this type, the frequencies with which

these plants bloomed differed between two groups of subjects. These differences allowed the creation of different abstract causal relations in these groups, so that *s* belonged to the abstract type that causes blooming for one group, but did not belong to the abstract type that causes blooming for the other group. Any difference between groups in their predictions regarding the outcome of the intervention must be due to the variation in this acquired abstract knowledge.

To test our proposal regarding maximal contrast, we presented stimuli that allow representation at various levels of abstraction (the features at the various levels being the candidate causes) and assigned maximal contrast to features at a certain level. To ensure that any influence of features at this level on causal judgment was a result of acquisition rather than perceptual salience, we chose features that were perceptually obscure at this level.

To disambiguate the nature of the induced causal knowledge, we controlled for the probability of blooming given the features of the stimuli (i.e., $P(\text{blooming}|c)$, where c is a feature). We did so because an obvious alternative explanation for our subjects' predictions is that the more *c* is associated with blooming, the more likely an item with *c* might be classified as tending to produce blooming. This criterion is consistent with the positive-test strategy, a prevalent hypothesis-testing strategy (Klayman & Ha, 1987). If ''blooming'' maps onto ''category'' and ''feature'' maps onto ''cue,'' then this criterion corresponds to what is called cue validity in the categorization literature (e.g., Reed, 1972; Rosch & Mervis, 1975; Rosch, 1978).

Hierarchical consistency predicts that the candidate *s* should be judged causal only by the group for whom the covariation involving *s* was hierarchically consistent. Because no alternative cause was available to explain the effect when the covariation was hierarchically inconsistent, however, subjects in this condition will lack confidence in their judgments. Because the set of substances presented, the cover story, and the cue validities for the candidate items were kept identical across groups, these variables cannot explain any observed difference in causal judgments between groups.

Method

Subjects

Ninety-six undergraduates of the University of California at Los Angeles participated in this experiment either to fulfill a course requirement or to earn 7 dollars.

Design

To manipulate hierarchical consistency as just described, we aimed to have two groups of subjects infer different abstract causal relations, each according to the maximal-contrast criterion. To assess the inferred level of abstraction, we examined whether subjects were sensitive to conditional contrast at, respectively, (1) the abstract level predicted by maximal contrast and (2) any level of abstraction. For this purpose, we constructed three *candidate* items (i.e., test substances that incorporated the target association) that were novel at different levels of

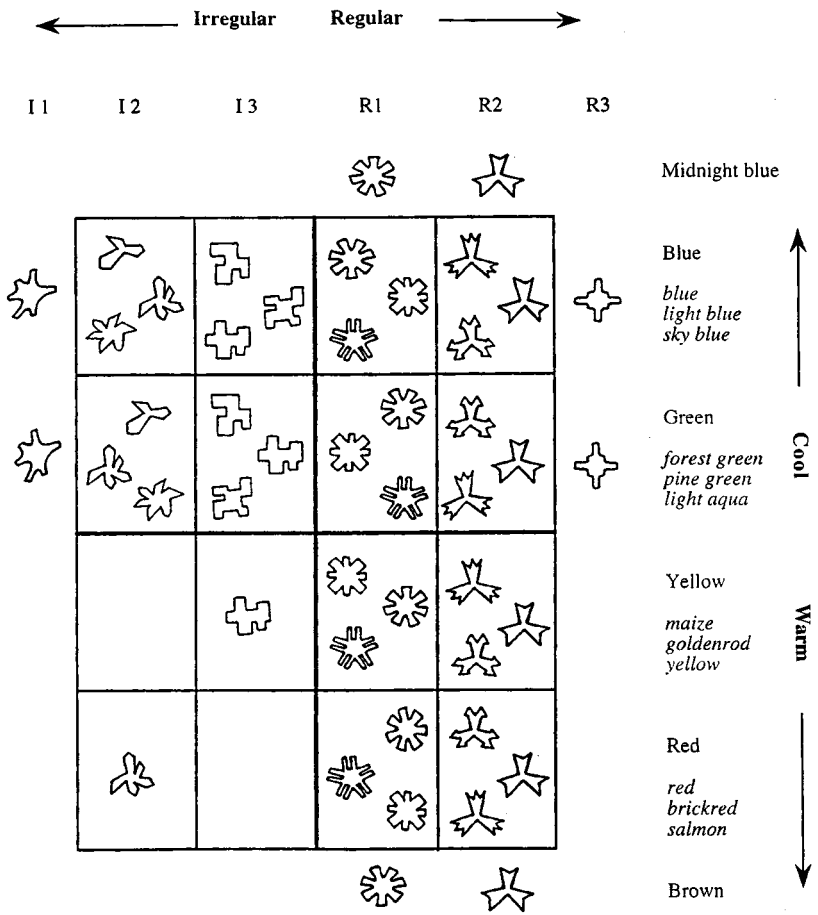


FIG. 1. Substances presented in the learning phase of Experiment 1.

abstraction. If subjects did infer a causal relation at the level of maximal contrast, their learning should generalize to an item that was novel at this level and they should judge hierarchical consistency accordingly. These candidate items were presented respectively to three subgroups of subjects within each of the two groups in the learning phase. In other words, the between-subject manipulation of consistency was replicated for three differently novel test stimuli. There were therefore six subgroups of subjects, with 16 subjects in each. Each subject was randomly assigned to one of the six groups.

Materials

In this section we describe characteristics of the entire set of stimulus items—substances fed to the plants. These stimuli varied systematically along the dimensions of color and shape. Except for the three candidate items, this set of stimuli was presented during the learning phase to every subject (see Fig. 1). (The color labels at the most specific level in the figure are Crayola crayon labels. The distinction between stimuli in the 4×4 region in the center and those in the surround are explained in the Procedure section.)

Levels of Abstraction

The stimuli allowed representation at at least three levels of abstraction along each of the dimensions of color and shape. As can be seen, with respect to color, an item can be represented as (1) a warm or a cool color at the most abstract of the three levels; (2) a type of color such as ''green'' at the middle level; and (3) a shade of a particular color, such as ''pine green.'' Analogously, the stimuli allowed encoding at at least three levels of abstraction with respect to shape. As can be seen, the shapes vary with respect to many features, the most important of which for our purpose is rotational symmetry. For brevity, we refer to rotationally symmetrical shapes as regular shapes and rotationally asymmetrical shapes as irregular shapes. An item could be represented as (1) an irregular or a regular shape, (2) a type of regular shape (e.g., R_2 in Fig. 1) or irregular shape (e.g., I_2), or (3) a variant of a particular type of regular or irregular shape.

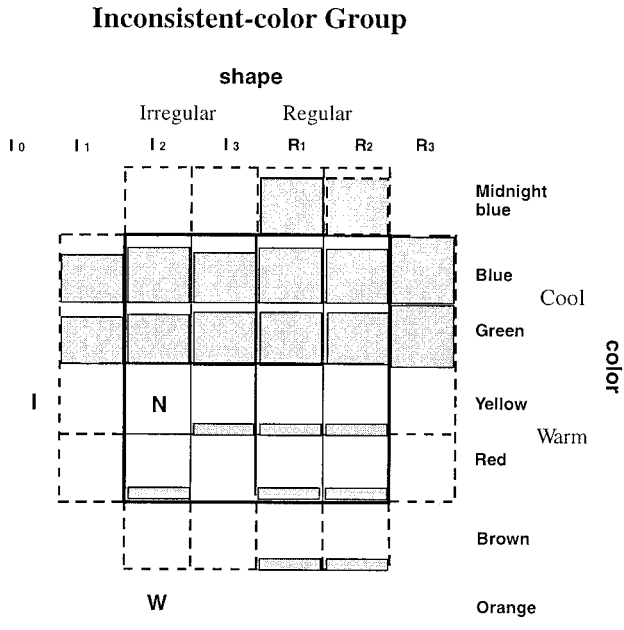
Consistency

Overview. As mentioned, to manipulate hierarchical consistency, we varied the frequencies of blooming for substances of various colors and shapes to allow subjects to infer different abstract causal relations between groups. In one group, shape was the causally relevant dimension (i.e., the one with a noticeable conditional contrast)—irregular substances produced a lot of blooming, whereas regular substances did not. In another group, color was the causally relevant dimension—cool-colored items produced a lot of blooming, whereas warm-colored items did not. The *candidate* items, presented in the test phase, were irregularly shaped and warm colored. Because the value of these items on shape (irregular) was the one that produced blooming for the former group, the covariation for this group between these items and blooming was hierarchically consistent; because the value of these items on color (warm) was the one that did not produce blooming for the latter group, the covariation for this group was hierarchically inconsistent. We therefore call the two groups the consistent-shape group and the inconsistent-color group.

Abstract causal relations in the knowledge base established in the learning phase. Figure 2 summarizes the frequencies of blooming (''blooming rates'' for short) for the various substances presented to the two groups. Labels on the axes in this figure correspond to those in Fig. 1 so that the blooming rates in Fig. 2 refer to stimulus items in the corresponding positions in Fig. 1. The proportion of area that is shaded within each color $\times$ shape cell represents the average proportion of plants that bloomed when fed with a substance of that color and shape. A ''0'' in a cell indicates that none of the plants fed with that type of substance bloomed. A cell with no shading or number indicates that that type of substance was not presented at all during the learning phase.

As can be seen in the figure, for the inconsistent-color group, most plants fed with cool-colored substances bloomed, but few plants fed with warm-colored substances did (see the horizontal partitioning of the diagram for this group as indicated by the proportions of shaded areas). In contrast, for the consistent-shape group, most plants fed with irregularly shaped substances bloomed, but few plants fed with regularly shaped substances did (see the vertical partitioning).

The appropriate *conditional* contrast for the dimension with high contrast is that based on all of the plants indicated in the figure (i.e., conditional on the environment in which they were grown). Because only one dimension for each group has a high contrast, no other dimension serves as an alternative plausible cause for conditionalization. For the inconsistent-color group, among such contrasts computed for values of color at various levels of abstraction (which were all relatively high), the maximum contrast was at the level partitioning warm from cool colors. To visualize the contrast for a value at a certain *level of abstraction,* see the difference between the average proportions of shading in Fig. 2 for that value (e.g., blue)



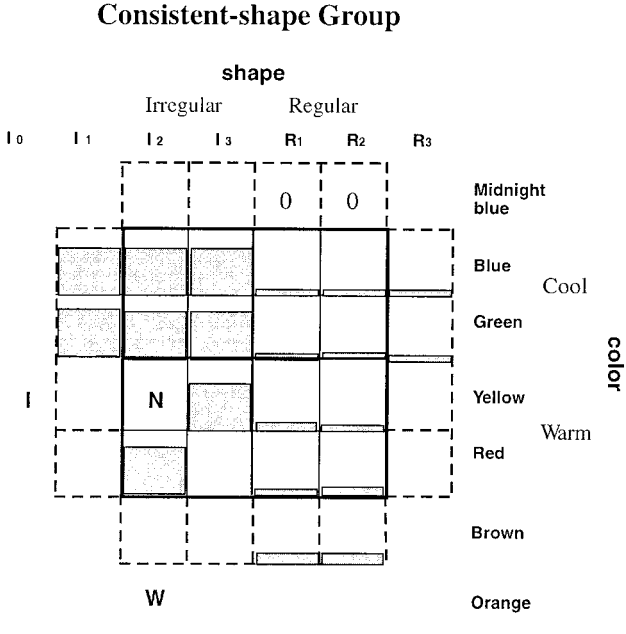


FIG. 2. The average proportion of plants that bloomed when fed with substances of various colors and shapes presented to the inconsistent-color and consistent-shape groups during the learning phase of Experiment 1.

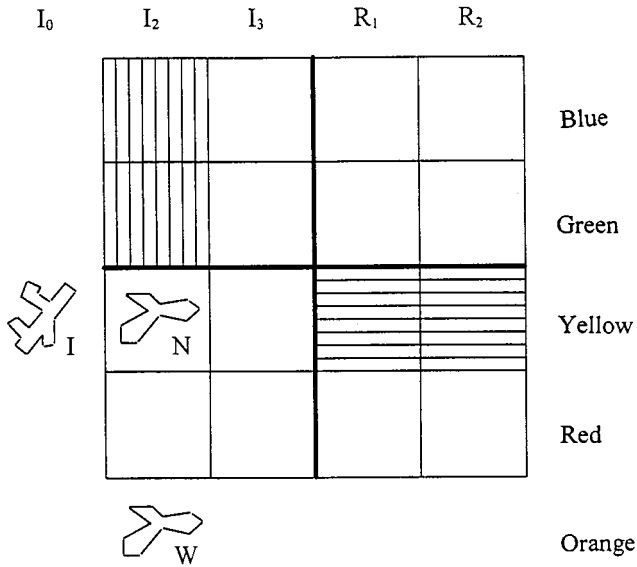


FIG. 3. Candidate items and their relation to the color-matched items (located in the area with horizontal stripes) and shape-matched items (located in the area with vertical stripes) in the categorization task of Experiment 1.

and for the rest of the values on that dimension (all nonblue colors). The mean conditional contrasts at each of four levels of abstraction (e.g., a shade of yellow, yellow, warm color, and all colors) were 0.46, 0.52, 0.67, and 0.44 respectively. Analogously, for the consistent-shape group, among the relatively high contrasts computed for values of shape at various levels of abstraction, the maximal contrast was at the level partitioning regular from irregular shapes. The conditional contrasts at each of four levels of abstraction for the consistent-shape group (e.g., a variant of shape I_2, shape I_2, irregular shapes, and all shapes) were 0.47, 0.53, 0.69, and 0.26, respectively.

It is logically possible that all values of the substances cause blooming, with some values causing blooming more than others. To rule out this possibility and ensure that warmth and regularity were unambiguously defined as causal features respectively for the two groups, with only one value at this level producing blooming, we matched the blooming rate of the unfed plants to the average ''low rate'' for that group: 2 in 10 for the inconsistent-color group and 1 in 10 for the consistent-shape group.

In summary, the dimension inferred to cause blooming should differ between groups: whereas the inconsistent-color group should infer that cool-colored substances cause blooming, the consistent-shape group should infer that irregularly shaped ones do so.

Candidate items covaried with the effect for both groups but varied in consistency between groups. Every candidate item had an irregular shape and a warm color. Figure 2 indicates the positions of the three candidate items relative to the knowledge base. These items are all located in the lower left quadrant and are denoted respectively by N, W, and I. (Pictures of these items appear in Fig. 3.)

As mentioned, the high blooming rate for a candidate item was presented to all subjects, creating a high covariation for both groups. As can be seen, however, the value of these items

on color (warm rather than cool) was the one that did not produce blooming for the inconsistent-color group (see the small shaded areas in the lower half of the knowledge base for this group), but the value of the candidate items on shape (irregular rather than regular) was the one that produced blooming for the consistent-shape group (see the large shaded areas in the left half of the knowledge base for this group).

Consistency at two levels of abstraction. Item I was of an old color, but had a novel type of *irregular* shape that did not belong to any type of such shapes presented during the learning phase. Accordingly, if causality was induced at the level of regularity, as our theory predicts, the covariation concerning item I would be consistent with the knowledge base for the consistent-shape group. But, if—contrary to our intended design—causality was not induced at a level as high as regularity, the covariation concerning item *I* would be inconsistent with the knowledge base for this group. For example, if a subject had induced that the causes of blooming were I_1, I_2, and I_3, an exhaustive list of less abstract known types of shapes that produce blooming, the covariation involving item I would be hierarchically inconsistent with any of these known relations. Item W was of an old shape, but had a novel *warm* color that did not correspond to any value of warmth presented during the learning phase. Thus, this item was inconsistent with the knowledge base for the inconsistent-color group at the level of warmth.

Item N instantiated a *novel combination* of a known warm color and a known irregular shape. The purpose of this item was to test whether subjects consider consistency with the causally relevant dimension (shape or color) regardless of the level of abstraction. Because the shape and color of this item were familiar, the covariation involving this item was consistent at all levels of abstraction for the consistent-shape group, but inconsistent at all levels for the inconsistent-color group.

Controlling for Cue Validity While Varying Contrast

Because the candidate items were irregular and warm, controlling for the cue validities of these items implies that both *P*(blooming|irregular) and *P*(blooming|warm) remain constant between groups. These two conditional probabilities are respectively estimated by the overall blooming rates in the left half and the bottom half of each diagram in Appendix A. This appendix lists the exact blooming rates given as the knowledge base for the two groups of subjects, where the rates refer to stimulus items in the corresponding positions in Figs. 1 and 2. Each number in the diagrams represents the number of plants that bloomed out of the 10 that were fed a given type of substance (the same number of plants were presented to the two groups). As simple arithmetic shows, each of the two conditional probabilities is equated across groups.

The outcome of the design was that while ''warm color'' was causal for one group and ''irregular shape'' was for the other, for *both* groups *P*(blooming|warm) was 0.2 and *P*(blooming|irregular) was 0.8. These probabilities were also the respective cue validities along these two dimensions for both groups at the next lower level of abstraction, the perceptually salient level. Thus, whereas hierarchical consistency predicts that the candidate items should be judged to cause blooming by the consistent-shape group only, the cue validities of these items predict that causal judgments should not differ between groups.

Procedure and Specific Materials

There were two phases—a learning phase and a test phase. The purpose of the learning phase was to establish the knowledge base about the substances and blooming. The purpose of the test phase was to assess subjects' just-acquired causal knowledge and the subsequent influence of this knowledge on causal judgments. No feedback was given during the test phase. Subjects were tested individually and allowed to spend as much time as they needed within a 1-h limit.

Learning Phase

At the beginning of the learning phase, subjects read the cover story described earlier.

Learning task. Forty-six types of substances (see Fig. 1) were presented during this task, which had two components. In the passive learning component, subjects were asked to examine information regarding the types of substances shown in the center 4 × 4 region in Fig. 1 (38 types). Information about each plant was presented on a card with a picture representing the substance fed to a plant (if it was fed) and a mark (a black dot) indicating whether that plant subsequently bloomed. These cards were arranged in rows of 10, with each row representing plants fed with the same substance. Subjects were asked to record how many out of each group of 10 plants bloomed. This information was available in both the learning and test phases.

In the dynamic learning component, subjects tested their hypotheses about what caused blooming. Subjects were asked to predict how many of 10 plants fed with various substances bloomed. Eight of the 10 items are shown outside the center 4 × 4 region in Fig. 1. The remaining two items were candidate items W and I. After each prediction, subjects were given the ''actual'' result for the eight substances that were not a candidate item. This feedback, listed in Appendix A, was not available during the test phase.

Test Phase: Testing the Role of Previously Acquired Knowledge in Causal Judgment

Then came the test phase. First, every subject was given a categorization task. This was followed by the presentation of the target association in which most plants fed with a candidate substance bloomed in the new environment, after which subjects were asked a question about the corresponding covariation and to give a causal rating for that type of substance.

Categorization task. The purpose of this task was to assess the level of abstraction of the causal relation established in the learning phase. Both groups of subjects were presented with a series of five sets of cards depicting various items. They were told that the gardener wanted to see if they understood what caused blooming. To reveal their understanding, they were to sort each set of substances into at most two groups by using the blooming information they had been just given. After each set, subjects justified why they sorted the items the way they did.

The task started with two practice sets that were analogous to the test sets except that they did not involve any candidate item. The three test sets, respectively containing one of the three candidate items, followed, with the ordering of these three sets counterbalanced across subjects. The three test sets each consisted of a candidate item (irregular and warm) and two old items: a known irregular and cool substance and a known regular and warm item. Therefore, the candidate item ''matched'' one old item only in shape and the other old item only in color. Figure 3 shows the three candidate items and their relation to the ''color-matched'' items and ''shape-matched'' items. For set N, the candidate item N was identical in shape to the ''shape-matched'' item and in color to the ''color-matched'' item. Being identical means sharing a value at all levels of abstraction of that value. The other two sets (set W and set I) were analogously constructed but with one difference: the value along one of the dimensions was shared only at an abstract level. For set W, item W matched the color of the ''color-matched'' item only at the level of warmth. For set I, item I matched the shape of the ''shape-matched'' item only at the level of regularity.

Recall that subjects were asked to sort according to what causes blooming. If subjects were sensitive to contrast at any level of abstraction at all, they should categorize items by the causal dimension. In particular, subjects in the inconsistent-color group should group item N with the color-matched item, whereas those in the consistent-shape group should group it with the shape-matched item. In addition, if subjects were sensitive to causality at the level of abstraction at which contrast was highest (e.g., regularity), those in the consistent-shape group

should group item I with the shape-matched item, even though that item had a different irregular shape as I, whereas the color-matched item was of an identical color as I. Similarly, subjects in the inconsistent-color group should group item W with the color-matched item with a different warm color, even though the third item had exactly the same shape as W. We constructed sets I and W so that these groupings would be unlikely unless subjects inferred causality at the respective levels of irregularity and warmth.

If subjects judged causality according to cue validity, both groups should group the candidate item with the shape-matched item; recall that for both groups, cue validity was higher for irregular shapes than for warm colors.

The target association. Subjects were told that a foreign friend obtained some plant seeds and a novel type of substance (items N, W, or I) from the gardener and raised the plants in his own yard in a distant country. They were then shown that 4 out of 5 of these plants subsequently bloomed. Recall that few unfed plants (2 of 10 for the inconsistent-color group and 1 of 10 for the consistent-shape group) in the gardener's yard bloomed. Because the location of the plants covaried perfectly with the novel type of substance (i.e., no information was available regarding plants grown in the new environment that were not fed with this or any other substances), the causal status of the candidate substance could not be determined by computing conditional contrast.

Causal rating. The purpose of this task was to evaluate the influence of hierarchical consistency. Causal judgment was measured by answers to a question involving a hypothetical intervention: ''Suppose the exact same substance were fed to a new group of plants in the gardener's yard. Would most of these plants have bloomed?''

Because only causal regularities license predictions regarding interventions, if the substance does not cause blooming, these plants would bloom at the rate of plants not fed any substance. Subjects answered this question using a rating scale ranging from −4 (''certain that most of the plants would not have bloomed'') to 4 (''certain that most of these plants would have bloomed''). A ''0'' meant that subjects had ''no idea'' whether the plants would have bloomed. They also wrote down the reason for their ratings.

Two practice trials involving old items preceded the critical trial involving the candidate item. In the learning phase, subjects in both groups had seen that most plants fed with the first practice substance bloomed, whereas most plants fed with the second practice substance did not.

It is possible that subjects might have failed to notice our instructions regarding the two environments. If they then unwarrantedly assessed the causal power of the candidate substance based on the perceived covariations, their predictions should not differ between groups. These covariations would lead to exactly the same estimated number of plants that would bloom for both groups: 4 of every 5.[2] Any difference in performance then must be due to the abstract causal knowledge acquired in the learning phase.

---

[2] Recall that Eq. (2) holds when causes alternative to c occur independently of it. Ignoring that the candidate substance covaried with the environment, and applying this formula unwarrantedly to estimate the causal power of the candidate substance (based on the blooming rates of the plants fed the substance in the friend's yard and of the unfed plants in the gardener's yard) would yield, respectively for the inconsistent-color and consistent-shape groups,

$$p_c = \frac{.8 - .2}{1 - .2} = \frac{3}{4} \text{ and } p_c = \frac{.8 - .1}{1 - .1} = \frac{7}{9}.$$

The estimated blooming rate of the hypothetical plants in the gardener's yard, assuming the same independencies, is $P(e|\bar{c}) + p_c - P(e|\bar{c}) \cdot p_c$. This is the probability of the union of two independent events: ''e produced by c'' and ''e produced by factors alternative to c.'' The probability of the latter event is estimated by $P(e|\bar{c})$—when c is absent, e is produced

Observation question. To confirm whether subjects observed a high covariation between the candidate substance and blooming, they were asked whether, compared to the percentage of unfed plants that bloomed in the gardener's place, a much greater percentage of the plants in the friend's place fed with the candidate substance bloomed. The ordering of the causal rating task and the observation question was counterbalanced across subjects.

At the end of the experiment, subjects were asked some background information, including whether they had taken a course that had a discussion of causality.

Predictions According to the Coherence Hypothesis and Competing Hypotheses

Hierarchical consistency predicts that all candidate items, regardless of the level of abstraction of their novelty, should be judged causal by the consistent-shape group but not the inconsistent-color group. It also predicts that the inconsistent-color group would lack confidence in their judgments, because no alternative cause explained the effect. The set of substances presented, prior causal knowledge evoked by the cover story, and the cue validities for the candidate items predict no difference in causal judgments between groups.

Finally, the base rate of blooming [i.e., *P*(blooming)] predicts exactly the opposite ordering of causal judgments as the coherence hypothesis. In the materials described, more plants bloomed for the inconsistent-color group than for the consistent-shape group (64% vs 36%); *P*(blooming) therefore predicts that the former group would be more likely than the latter group to judge that a novel item produces blooming.

Results

Categorization Task

Categorization performance confirmed that the two groups of subjects indeed induced the causal dimension. Moreover, they did so at the level with maximal contrast. Table 1 shows the percentage of subjects in the two groups who categorized the various candidate items with the color-matched item, the shape-matched item, or neither. As can be seen, most subjects in the inconsistent-color group categorized the candidate items by color, whereas most subjects in the consistent-shape group categorized them by shape, $p < .001$ for each comparison by Fisher's exact test.

Observation and Causal Rating Tasks

The ordering of these tasks had no effect; we therefore report results from here on collapsing over orderings. All subjects answered the observation question correctly, confirming that they uniformly perceived a high covariation between the candidate substance and blooming.

Recall that we defined maximal contrast for features that were nonobvious.

---

by only factors alternative to *c.* Thus, the estimated blooming rate for the hypothetical plants would be $.2 + 3/4 - .2 \times 3/4$ for the inconsistent-color group and $.1 + 7/9 - .1 \times 7/9$ for the consistent-shape group, both being equal to .80, as should accord with simple intuition. If subjects ignored information about other substances and assumed that the environment was irrelevant, plants fed with the candidate substance should be estimated to bloom at the same rate in both yards.

TABLE 1
Percentages of Subjects in the Inconsistent-Color and Consistent-Shape Groups (n = 48 in each) Who Grouped Particular Candidate Items with the Color-Matched Item (C-M), the Shape-Matched Item (S-M), or Neither in the Categorization Task of Experiment 1

	Item grouped with candidate item		
Group	C-M	S-M	Neither
		Item N	
Inconsistent-color	73	23	4
Consistent-shape	10	88	2
		Item W	
Inconsistent-color	56	42	2
Consistent-shape	2	94	4
		Item I	
Inconsistent-color	88	6	6
Consistent-shape	25	69	6

Given the time constraints in our experiment, it is likely that some subjects never discovered the relevant but nonobvious abstract representation of the stimuli, in which case they could not have computed contrast with respect to it. Because a prerequisite for testing the influence of hierarchical consistency is that subjects did possess the abstract causal knowledge by which consistency is defined, we restrict the evaluation of such influence to those subjects who did infer the intended causal knowledge. Specifically, the analysis of the causal ratings for each candidate item was restricted to subjects who had categorized that item according to the causal dimension (79% of subjects in the consistent-shape group and 71% of those in the inconsistent-color group). Our restriction of the sample provides a more informative test of coherence due to hierarchical consistency per se, unconfounded by subjects' ability to learn the relevant causal feature.

Did the Consistent-Shape Group Rate the Candidates as More Causal?

As predicted by the coherence hypothesis alone, subjects judged a covariation to be more causal (i.e., more positive) when it was hierarchically consistent than when it was not. Figure 4 shows the mean causal ratings for each candidate item in the two groups (for subjects who categorized the respective item according to the causal dimension). The difference between groups in the mean rating for each item is highly reliable, $t(22) = 7.8$ for N, $t(22) = 4.5$ for W, and $t(22) = 3.8$ for I, $p < .001$ for each comparison. The result for item N supports our hypothesis that causal judgments are based on consistency with respect to contrast, at least at some level of abstraction. Those

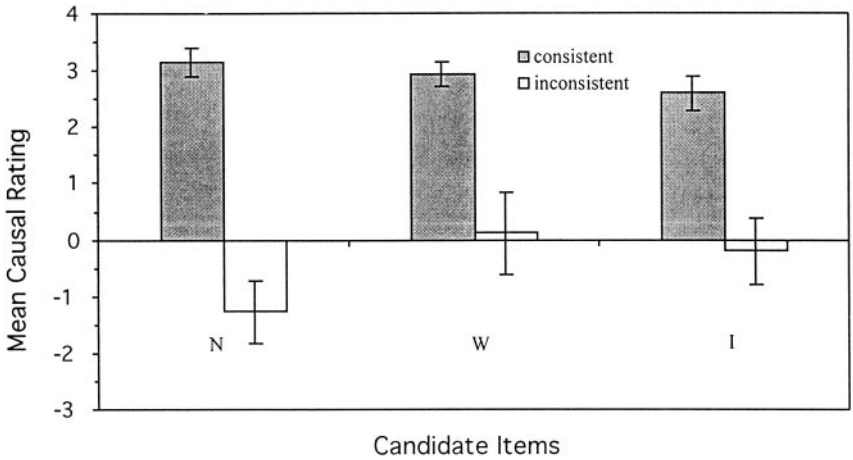


FIG. 4. Mean causal ratings for the candidate items given by subjects in the inconsistent-color and consistent-shape groups in Experiment 1 who sorted the respective item according to the relevant causal dimension in the categorization task.

for items W and I provide evidence that subjects defined consistency at the level of maximal contrast.

Was the Consistent-Shape Group More Confident Than the Inconsistent-Color Group?

Recall that subjects in the inconsistent-color group were presented with a situation in which there was no known or highly plausible cause that serves to explain the target effect. As predicted by the coherence hypothesis, these subjects were unsure of their judgment. Only the mean rating for N was reliably below 0, $t(10) = 2.35$, $p < .05$; those for W and I did not reliably differ from 0, $t(8) = 0.15$ for W and $t(13) = 0.37$ for I, $p > .5$ for each comparison. In contrast, subjects in the consistent-shape group, who should not have any unexplained regularity, were quite confident of their causal judgments. The mean rating for every candidate item was reliably above 0, $t(12) = 12.6$ for N, $t(14) = 14.0$ for W, and $t(9) = 8.67$ for I, $p < .001$ for each comparison.

To directly compare the certainty indicated by the two groups, we consider the *absolute* values of the *mean* ratings for each item. As predicted by the coherence hypothesis, these absolute values were reliably lower for the inconsistent-color group than for the consistent-shape group: $t(22) = 3.3$ for N, $t(22) = 4.5$ for W, and $t(22) = 3.2$ for I, $p < .01$ for each comparison, indicating that subjects in the inconsistent-color group were less confident about the causal status of the candidate items.

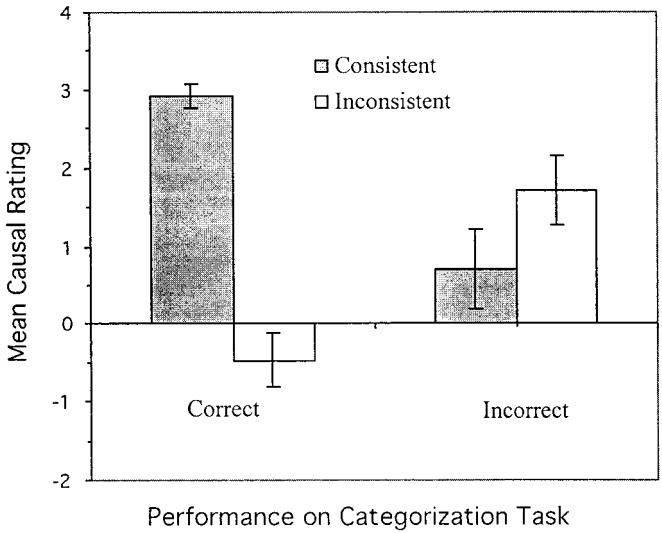


FIG. 5. Mean causal ratings for the candidate items given by subjects in the inconsistent-color and consistent-shape groups of Experiment 1 who respectively did and did not sort the items according to the relevant causal dimension in the categorization task.

Does Performance on the Categorization Task Predict Causal Rating?

To provide further evidence that it was causal knowledge successfully inferred during the learning phase that produced the just reported differences in causal judgments regarding the candidate items, we analyzed whether categorization performance predicts causal rating. Because there were too few subjects who sorted an item incorrectly (5, 7, and 2 respectively for items N, W, and I) for a statistical analysis of performance on each item alone, we collapsed over all subjects who gave causal ratings on an item that had been sorted incorrectly into one group and compared their causal ratings of these items with those given by subjects who had sorted those items correctly. As shown in Fig. 5, for the inconsistent-color group, the incorrect subjects, compared to the correct subjects, rated the candidate items as more causal, $t(df = 46) = 3.52$, $p < .001$. For the consistent-shape group, the incorrect subjects, compared to the correct subjects, rated the candidate items as less causal, $t(df = 10.6$ due to unequal estimated population variances$) = 4.12$, $p < .002$. The mean causal ratings for the incorrect subjects did not differ reliably between the consistent-shape and inconsistent-color groups, $t(df = 22) = 1.50$, n.s., with the difference being in the opposite direction as that observed for the correct subjects. This pattern of results is exactly what would be expected if the level of abstraction of the acquired causal relation determines hierarchical consistency.

Discussion

We proposed that when there is insufficient information for computing contrast for a candidate cause while controlling for alternative causes so that a covariation between the candidate and the effect does not in itself imply causality, causal judgments regarding that covariation will depend on how the reasoner's existing system of causal rules explains or does not explain the covariation. Experiment 1 manipulated explanation by causal rules that were induced from observable frequencies in a situation in which no alternative cause explains the effect. As predicted by our coherence hypothesis, when a covariation was hierarchically consistent, reasoners judged it as genuinely causal; when this relation was hierarchically inconsistent and the effect was not explained by an alternative cause, reasoners judged it as spurious, but they were less confident of their judgment. This observed difference suggests that the traditional views need to be amended with assumptions about coherence. Our results carry two additional implications. First, they demonstrate that genuine-versus-spurious causal judgments regarding a covariation can be explained by superordinate causal knowledge inferred on the basis of observable information. Second, the scope of the cause with respect to a given effect is defined by the level of abstraction at which its contrast is maximum.

Our experiment ruled out several other explanations. The set of stimuli presented, frequency information regarding a candidate item itself and the cue validities of its features, and the prior causal knowledge subjects brought to the experiment, which were kept constant between the inconsistent-color and consistent-shape groups, cannot explain the systematic differences in performance obtained between these groups. Neither can the base rate of blooming, which predicts results opposite to that observed.

Supplementary Tests of Alternative Explanations

The evaluation of three alternative interpretations of our results requires additional evidence. We discuss and refute these interpretations below.

Were Regularity and Warmth Salient Features of Shape and Color Respectively?

It is possible that, despite the intention underlying the construction of our stimuli, warmth of color is a perceptually salient *feature* of color and regularity of shape is a perceptually salient *feature* of shape. If so, a variant of the conditional-contrast explanation is plausible: once subjects had learned which dimension (color or shape) was causal according to conditional contrast, they would spontaneously sort items based on warmth or regularity. To show that features at these levels of abstraction were not spontaneously salient, but rather induced as a result of maximal contrast, we conducted a clustering task outside of the causal context. In this task, a separate sample

of 20 undergraduates from the same subject pool were asked to sort cards depicting *all* 49 types of items used in Experiment 1 into groups. They were asked to ''sort these items into groups based on similarity in shape (or color),'' in the way that they felt was ''most natural'' to them. We allowed subjects opportunities to do so over multiple trials rather than a single trial, to provide a more sensitive measure of how ''unnatural'' were the features at the level of maximal contrast. After the first trial, they were asked after each sorting, ''If you have to further sort these items into fewer groups based on similarity in shape (or color), how would you group them?'' On the sixth trial, if they had not sorted the items into two groups, they were explicitly told to do so. Then they were asked to sort the items according to the alternative dimension. The ordering of shape and color was counterbalanced across subjects.

The results show that regularity and warmth were not the most perceptually salient features to any of our subjects. None of the 20 subjects grouped the items according to either regularity or warmth on the first trial. When subjects sorted by shape, only one sorted the items by regularity, and she did so only on the final trial. When they sorted by color, no one sorted by warmth earlier than the third trial. In total, 12 subjects separated warm from cool colors by their final trial; however, all of these subjects did so only after several other ways of sorting, presumably according to more perceptually salient features.[3]

Corroborating our results in the clustering task, none of the 96 subjects in Experiment 1 ever justified their categorization or causal-rating performance by similarity to either of the warm and irregular items, even though these items could be readily referred to, as they remained in view through out the experiment. Instead, most of them (93%) justified their performance by referring to a feature of either color or shape, consistent with learning based on conditional contrast.

[3] These results also refute all other explanations based on perceptual salience. For example, it might be argued that subjects used conditional contrast to merely infer a collection of causal relations at a more specific level (e.g., that items with an I_1-type shape cause blooming, items with an I_2-type shape cause blooming, and so on, but items with an R_1-type shape do not cause blooming, etc.); however, they then generalize to *similar* values along the causal dimension, thus presumably explaining performance on the candidate items. This argument rests on a critical implicit assumption: the feature that defines similarity for the generalization (regularity or warmth for our materials) is salient. Otherwise, out of the many potential features of a dimension on which similarity could be measured (similar in shape due to having pointy tips, similar in shape due to having the same thickness of the outline, etc.), why should a particular feature and only that feature (regularity) count? Another variant of the perceptual similarity argument is that subjects simply rated the candidate items by assuming that they should have effects similar to those of the most perceptually similar items, in particular, the other two items that were irregularly shaped and warm colored (see Fig. 1). If so, subjects in the consistent-shape and inconsistent-color groups (in Experiment 1) would *not* have sorted the test items differently, as they clearly did: the other two items in the categorization sets were *a priori* just as similar to a candidate item in one group as in the other group.

Can the Difference in Covariation Explain the Difference in Causal Ratings?

To equate the cue validities of the candidate items while ruling out an interaction between color and shape, we assigned a slightly higher blooming rate in the inconsistent-color group than in the consistent-shape group to items that were irregular or warm but not both. This led to a slightly higher covariation for a candidate item in the consistent-shape group than in the inconsistent-color group. If the observed difference between groups in causal ratings was due to this variation in covariation, a similar difference should be observed between subjects presented respectively with these covariations even when the abstract causal knowledge that establishes hierarchical consistency is omitted. To test this argument, we added two conditions that measured causal ratings on candidate items with the respective covariations but without that abstract knowledge.

Seventy subjects were randomly assigned to two conditions, in each of which they were given a cover story and a causal rating task highly similar to those in Experiment 1, but the learning phase was omitted so that the abstract causal knowledge inferred in Experiment 1 could not have been inferred here. (The categorization task, for which no feedback was given, was also omitted.) These subjects were a separate sample from the same pool as the previous studies. They were asked the exact same hypothetical questions as those in Experiment 1 about the same number of test substances, labeled as A, B, and C in this context (no pictorial stimuli were presented). As before, they were told that items A and B (practice items) had been separately fed to a group of 10 plants in the gardener's place and item C (the one corresponding to the candidate item in Experiment 1) had been fed to a group of 10 plants in his friend's backyard in a foreign country. In addition, 10 plants were raised in the gardener's place without being fed any substance. For one of these conditions, which we call the ''color-corresponding'' condition, the frequencies of blooming for these four groups of plants matched the corresponding frequencies given to the inconsistent-color group, whereas for the other condition, which we call the ''shape-corresponding'' condition, they matched the corresponding frequencies in the consistent-shape group.

Our results indicate that the difference between groups in the covariation for a candidate item cannot explain the observed differences between the inconsistent-color and consistent-shape groups in their causal ratings of that item. The mean rating for item C did not differ between the color-corresponding and shape-corresponding conditions: it was, respectively, $1.50 \pm .25$ ($n = 36$) and $1.65 \pm .28$ ($n = 34$) in these two conditions, $t\ (39) = 0.39$, $p > .5$. As would be expected if the hierarchical consistency established in Experiment 1 influenced causal ratings, the mean rating for item C in the color-corresponding condition was significantly higher than the mean rating for W reported earlier (0.11) for the inconsistent-color group of Experiment

1, $t(43) = 2.25$, $p < .03$. Likewise, the mean rating for item C in the shape-corresponding condition was significantly lower than the mean rating for I reported earlier (2.60) for the consistent-shape group, $t(42) = 1.72$, $p < .05$ (one-tailed). These results show that hierarchical consistency established by the knowledge base, rather than the small difference in covariation for a candidate item, explains the difference in causal ratings between the inconsistent-color and consistent-shape groups.

Did Academic Training Influence the Distinction between Genuine and Spurious Causes

One might argue that although the coherence hypothesis is supported by our results, it does not describe what ordinary people do because the academic training received by our subjects—university students—taught them to use conditional contrasts in assessing causality. This hypothesis is implausible judging from previous research indicating a close parallel between human causal inference and Pavlovian conditioning in species that do not receive any academic training (e.g., Cheng, 1997; Shanks & Dickinson, 1987). Nonetheless, to directly assess the influence of academic training on the use of conditional contrasts, we further analyzed the categorization results of Experiment 1 with respect to whether subjects had taken any courses involving a discussion of causality.

Of the 96 subjects, only 25% thought that they had taken one or more relevant courses. Introductory psychology was mentioned six times, statistics and philosophy each five times, research methods four times, and other courses less often.

The percentage of subjects who categorized according to the causal dimension ranges from .79 to .83 for those who had taken a relevant course, and from .74 to .81 for those who had not. There is no significant difference in sorting performance between subjects who had taken the courses and those who had not, $\chi^2\,(1) = 0.02$ for N, $\chi^2\,(1) = 0.07$ for W, and $\chi^2\,(1) = 0.18$ for I, n.s. for each comparison.

In sum, most of our subjects did not report having been taught about causality. Moreover, reporting having taken a relevant course did not influence how well they induced the causal dimension.

EXPERIMENT 2: EXPLAINED SPURIOUS CAUSAL RELATIONS

According to our coherence hypothesis, when an alternative cause explains a covariation, this alternative cause should increase coherence and hence the confidence accompanying the judgment that the candidate is a spurious cause. The purpose of the present experiment is to extend the generality of the coherence hypothesis to this kind of situation. Recall that in both conditions of Bullock's (1979; Bullock et al., 1982) Jack-in-the-box experiment, the spurious cause occurred in a situation in which a highly

plausible co-occurring cause could explain the effect. In the single-box condition, the traveling light was a spurious cause, but the popping of Jack was explained by the rolling ball, the genuine cause. In the separate-box condition, the roles of these candidates were reversed. Bullock et al.'s pattern of results has been widely interpreted as evidence contradicting the covariation view. Our goal was to obtain this pattern of results by a change in hierarchical consistency.

To explain her pattern of results, we created an analogous situation in which a spurious cause and a co-occurring genuine cause reversed their roles between conditions. Instead of manipulating unspecified prior causal knowledge (by the separation of the boxes in her experiment), however, we manipulated observable frequencies to create a different knowledge base in two groups of subjects, thereby pinpointing the relevant superordinate causal knowledge as we did in Experiment 1.

As in Experiment 1, subjects were asked to imagine that they were applying for a job as a gardener's assistant. Their primary task was to evaluate the causal status of each of several candidate substances that covaried with blooming, and the contrast for these substances could not be computed while controlling for alternative causes because of a change in the environment, from the gardener's yard in the learning phase to the foreign friend's yard in the test phase. As before, we varied the observable frequencies presented during the learning phase so that one group of subjects should infer that color was causal, but the other group should infer that shape was. In contrast to the previous design, however, *two simultaneously presented candidate items, rather than one, covaried with the effect,* the blooming of the plants to which they were fed. The foreign friend obtained several pairs of substances from the gardener, along with the plant seeds, and raised the plants in his own yard in a distant country, feeding each group of plants a pair of substances simulataneously. The items in one of these candidate pairs had novel colors; those in the other pair had novel shapes. As before, novelty on each dimension measures the level of abstraction of the respective inferred causal relation. For each of the two pairs of co-occurring items, the covariation involving one item was hierarchically consistent with prior causal knowledge for one group and inconsistent for the other group; the hierarchical consistency of the *other* item in the pair was reversed across the two groups.

Given the difference in hierarchical consistency for the items in a pair and the reversal of the roles of the items between groups, our coherence hypothesis predicts that whereas one item in a pair would be judged as *the* cause by one group, the other item would be judged so by the other group. Such a pattern of results would show that an interaction between candidate causes and prior causal knowledge (e.g., Bullock et al., 1982) can be obtained by manipulating hierarchical consistency. Finally, because (unlike in Experiment 1) the highly plausible co-occurring cause explains the effect, leaving

no residual inconsistencies, the coherence hypothesis also predicts that subjects will identify a hierarchically inconsistent candidate as a spurious cause with confidence.

Similar to Experiment 1, a learning phase with a passive and a dynamic component was followed by a test phase, in which subjects were given a categorization task, the target associations, an observation question about the corresponding covariations and a causal rating task. Unlike Experiment 1, which varied candidate items between subjects, the present experiment varied pairs of candidate items within subjects.

Method

Subjects

Fifty-six undergraduates of the University of California at Los Angeles participated in this experiment to fulfill a course requirement. Subjects were randomly assigned to one of two groups. There were 25 subjects in one group and 31 in the other.

Design

The design of the present experiment shared three constraints with that of Experiment 1: (1) the stimuli could be represented at multiple levels of abstraction, (2) the causal dimension in the knowledge base was varied between two groups of subjects, and (3) cue validity was controlled (in the case of the present experiment, the cue validities of one candidate item in each pair were held constant between groups). But, unlike Experiment 1, the present experiment manipulated hierarchical consistency in the context of an alternative candidate that explains the effect.

Manipulating Hierarchical Consistency

All subjects were presented with an identical association regarding the pairs of candidate items. Prior to evaluating the substances, each subject was presented observable frequencies regarding the blooming of plants and the various types of substances fed to the plants, which provided sufficient information to allow causal inferences. This knowledge base was manipulated between two groups of subjects to yield the patterns of hierarchical consistency for the pairs of candidate items. Our primary dependent measure was an indirect causal assessment involving a recommendation for each item in these pairs.

To specify how we manipulated hierarchical consistency for the candidate items in each pair, Fig. 6 shows the pattern of blooming frequencies in the knowledge base in relation to the candidate items for the two groups of subjects (see Appendix B for the corresponding exact frequencies; we explain items N and N_p later). As in Experiment 1, a different dimension of the stimulus items causes blooming for each group: color for one group (with contrast

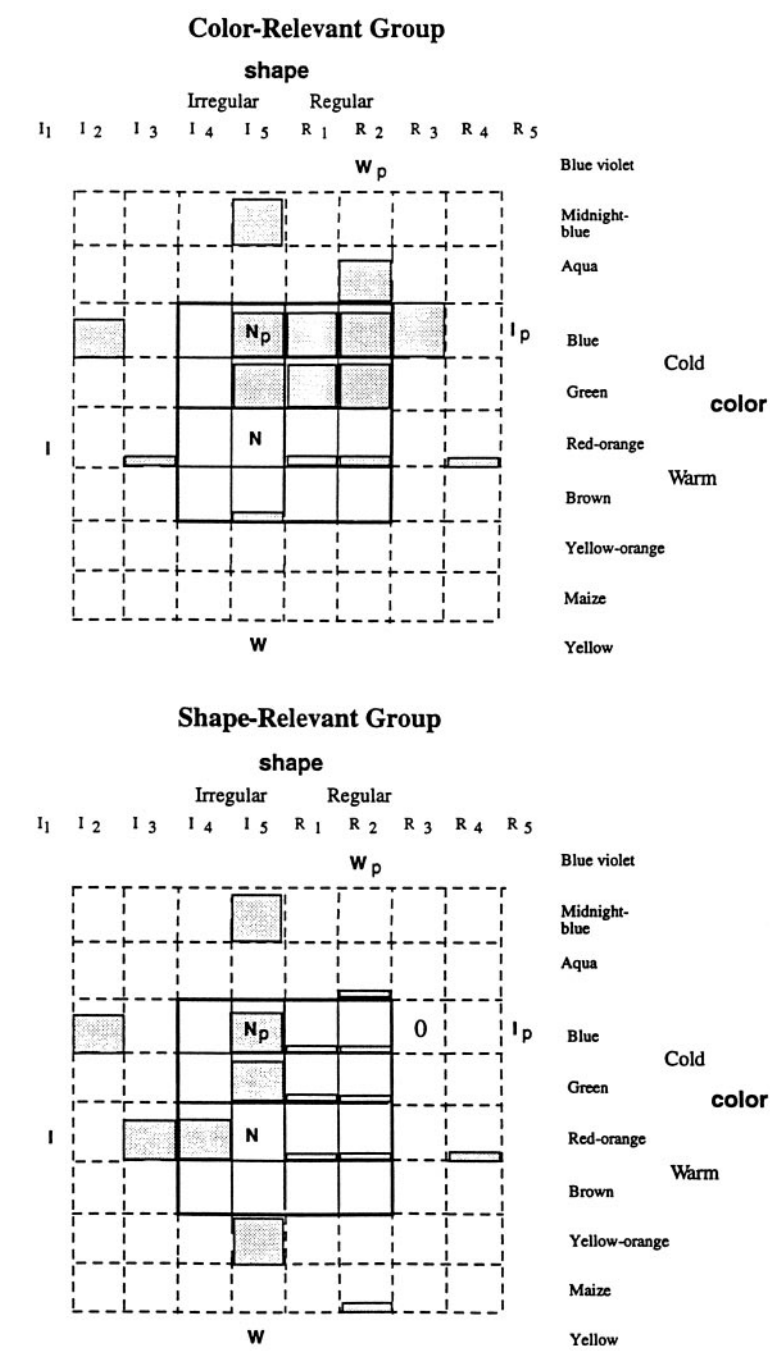


FIG. 6. The average proportion of plants that bloomed when fed with substances of various colors and shapes presented to the color-relevant and shape-relevant groups in the learning phase of Experiment 2.

being maximal at the level of warmth) and shape for the other group (with contrast being maximal at the level of regularity). For each of the two pairs of candidate items that tested the role of hierarchical consistency (W-W_p and I- I_p in the figure), one item had a warm color and an irregular shape (W and I), but the other had a cold color and a regular shape (W_p and I_p). The subscript ''p'' represents a ''paired item''. The pair W-W_p had novel values along the dimension of color, whereas I-I_p had novel values along the dimension of shape (Fig. 7 shows pictures of these two pairs of test substances; Fig. 8 shows pictures of the stimulus items presented in the learning phase). Warm colors, and only warm colors, were associated with little blooming according to the *color-relevant* knowledge base, but irregular shapes, and only irregular shapes, were associated with much blooming according to the *shape-relevant* one. Therefore, the covariation regarding items W and I were inconsistent with the color-relevant knowledge base, but consistent with the shape-relevant one. In contrast, the covariations regarding their paired items W_p and I_p were consistent with the color-relevant knowledge base, but inconsistent with the shape-relevant one. Accordingly, the coherence hypothesis predicts that, whereas the covariation involving candidate items W and I would be interpreted as spuriously causal by the color-relevant group but as genuinely causal by the shape-relevant one, the corresponding covariation involving candidate items W_p and I_p would be interpreted the opposite way by the two groups.

Controlling for Cue Validities

Given the constraints of our design, it was impossible to control for the cue validites of both items in a pair between groups. We therefore controlled for the cue validities of only one item in each pair. The cue validities of candidate items W and I were kept exactly constant across the two groups for the categorization task, as indicated by the blooming rates listed within the 4×4 center region in Appendix B (only rates within this region were presented prior to the categorization task). If one includes the blooming rates of the items outside this region, however, the cue validities of the candidate items were kept only approximately equal for the causal rating task (during which the former but not the latter set of rates was available). The changes in cue validities due to these extra frequencies were minor, however, and do not predict a *shift* in the causal dimension (i.e., the causal ratings for the candidate items should still be positive for both groups). Including the frequencies of these novel items would decrease the cue validity of irregularity from 0.80 to 0.76 in the color-relevant group, but increase it from 0.80 to 0.81 in the shape-relevant group. It would also increase the cue validity of warmth from 0.20 to 0.21 in the color-relevant group and from 0.20 to 0.32 in the shape-relevant group. Notice that for both groups, the cue validity of irregularity (ranging from 0.76 to 0.81) would still be much higher than that of warmth (ranging from 0.20 to 0.32). Therefore, if subjects rated the

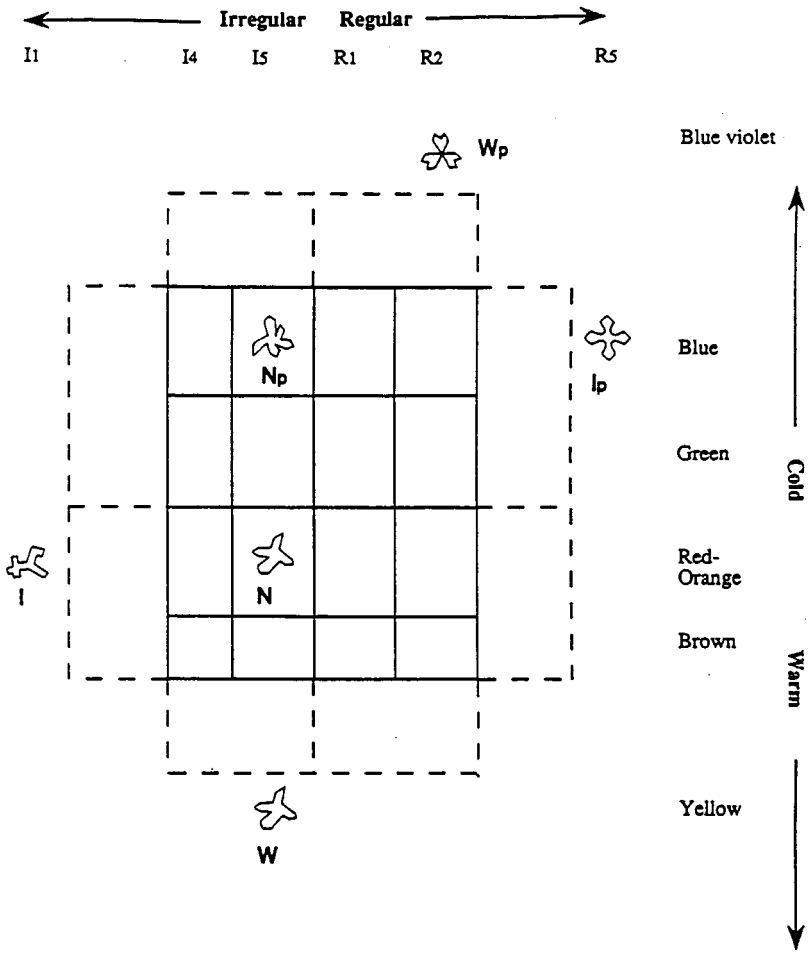


FIG. 7. Candidate items in Experiment 2.

causality of these candidate items according to cue validities, both groups would still rate these items according to their irregular shape, although the shape-relevant group might be more likely than the color-relevant group to judge that the candidate items cause blooming.

The base rate remains much higher for the color-relevant group than for the shape-relevant group, yielding predictions opposite to those based on hierarchical consistency.

Ruling out Cue Validity as an Alternative Explanation: Candidate Pair N-N_p

In addition to the two pairs of candidate items described earlier, we created a third pair of candidate items to rule out an alternative explanation due to

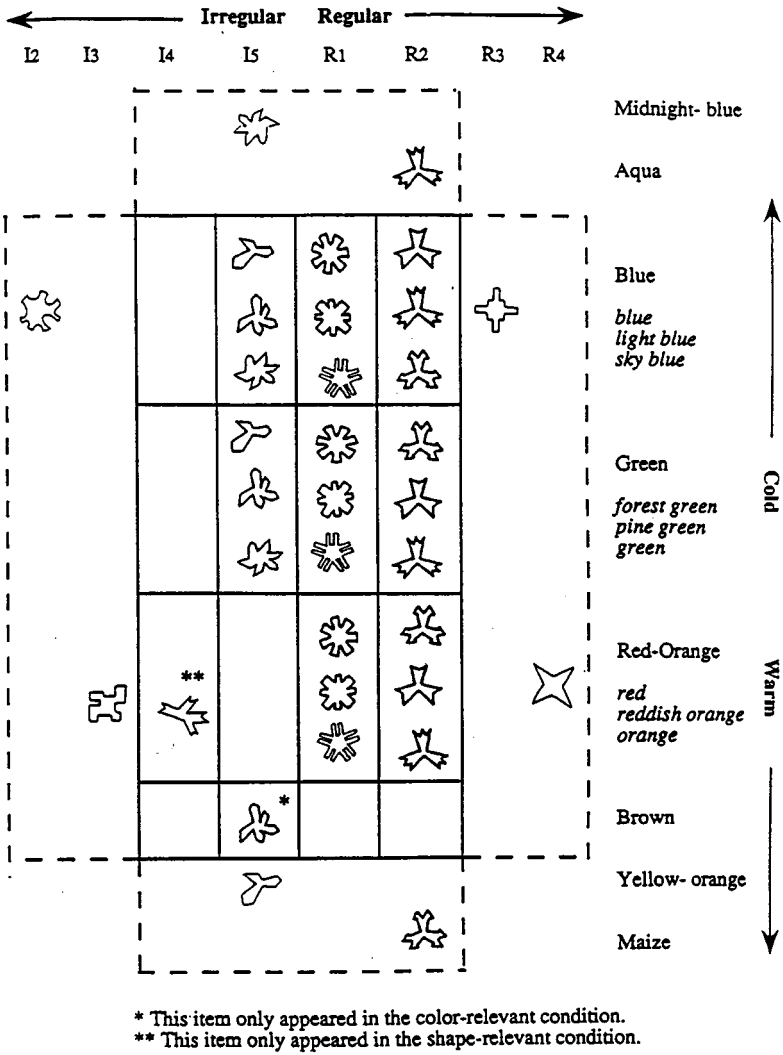


FIG. 8. Substances presented in the learning phase of Experiment 2.

variations in cue validity. Recall that we controlled for the cue validities of only one item in each pair, allowing that of the other item in the pair to vary between groups as a result of adopting an opposite value on hierarchical consistency, as required by our manipulation of consistency. Thus, the cue validities of W_p and I_p were left to be high for the color-relevant group (e.g., the cue validity for being regular were 0.66 and that for being cold was 0.90) and low for the shape-relevant group (e.g., the cue validity for being regular was 0.10. and that for being cold was 0.35). It might therefore be argued that these variations can explain the causal ratings for these uncontrolled

items directly and those for the controlled items indirectly. For example, consider pair *W*-W_p (see Fig. 6): the higher cue validities for W_p for the color-relevant group than for the shape-relevant group predicts that the former group would give W_p a higher causal rating (than would the latter group); now, if one allows the additional assumption that *subjects judged one and only item in each pair to be causal,* then the color-relevant group would give *W* a lower causal rating than would the shape-relevant group. This pattern of ratings coincides with that predicted by the coherence hypothesis for both the controlled and uncontrolled item.

The purpose of the third pair of candidate items, *N* and N_p in Figs. 6 and 7, was to rule out this explanation due to cue validity by refuting the assumption that only one item in a pair will be judged causal. Unlike the items in the earlier two pairs, which were both novel, this pair consisted of a novel substance, *N,* and a known substance, N_p. As for items *W* and *I,* the cue validity of item *N* was kept the same across groups. In the learning phase, subjects in both groups were shown that plants fed with N_p bloomed more than the unfed plants. This known substance could therefore be inferred to be causal from its contrast conditional on the old environment. As for the earlier two pairs of candidate items, this substance and the novel one in combination, as a candidate pair, were then associated with much blooming in a different environment. Under the same assumption of the cue-validity explanation as for the other item pairs (that subjects judge one and only item in each pair to be causal), because N_p should be rated causal by both groups, *N* (the novel item in the pair) should be rated noncausal by both groups.

The presentation of this pair of items is analogous to the traditional blocking paradigm in which a well-established CS (the known item) and a novel CS (the novel item) jointly predict the presence of a US (blooming). Differing from that paradigm, however, the conditional contrast for the novel item (the novel CS) could not be computed in our experiment (because of the change of environments). Current models that explain ''blocking'' (e.g., Rescorla & Wagner, 1972), which by default ignore the possibility of representation at various levels of abstraction, predict that the novel item in the pair should be rated noncausal by both groups.

In contrast to both of these accounts, the coherence hypothesis predicts that covariation involving the novel item should be judged noncausal only if it is hierarchically inconsistent. Otherwise (if the covariation is consistent), despite the fact that a ''competing'' cause co-occurs with the candidate, the candidate would be judged causal (i.e., there will be no blocking)—both items in the pair would be judged as genuine causes. To test these predictions, we manipulated the hierarchical consistency concerning *N:* As for items *W* and *I,* the covariation concerning this item was consistent with the shape-relevant knowledge base and inconsistent with the color-relevant one. The covariation concerning the known item N_p, however, was consistent with the knowledge base in *both* groups. The coherence hypothesis therefore predicts

that *N* would be judged as a spurious cause (i.e., its learning would be blocked by the known cause N_p) only in the color-relevant condition.

Materials and Procedure

The present experiment used the same materials and procedure as in Experiment 1 except for the differences mentioned in the following sections. Subjects were allowed an hour to complete the experiment at their own pace.

Learning Phase

The present experiment presented fewer items than did Experiment 1 to reduce the complexity of the knowledge base and the time required to learn it.

The center 4 × 4 region of Fig. 8 demarcated by solid lines shows the substances presented in the passive learning component. As this figure shows, one item presented in this component for each group was warm-colored and irregularly-shaped (see the items marked by asterisks); we presented their blooming rates to eliminate the possibility of an interaction between shape and color. Although these two items differed between groups, we show that this difference cannot explain the difference in causal judgments between groups.

The eight novel substances outside the center 4 × 4 region in Fig. 8 were presented during the dynamic learning component in which the task was to *predict* whether plants fed with each of various items bloomed. The feedback for these items (the numbers enclosed in parentheses in Appendix B) was given immediately *after* subjects have completed the categorization task (the next task).

Test Phase

Categorization task. As in Experiment 1, in addition to evaluating the computation of contrast, the categorization task was used as a pretest to determine whether the manipulation of consistency was successful. The version here began with a practice set followed by three test sets, given in the order *N, W,* then *I.* Each set consisted of cards respectively depicting a candidate item and two other items, and was analogous to the correspondingly labelled set in Experiment 1.

Set N consisted of candidate item *N,* which exhibited a novel variant of a known type of irregular shape and a novel shade of a known type of color (slightly different from the item with the same label in Experiment 1). One of the remaining items in this set was identical to *N* in shape and the other was identical to it in color. In set W, the color-matched item only shared the value of color with candidate item *W* at the relatively abstract level of warmth. Analogously for Set I, the shape-matched item only shared the value of shape with candidate item *I* at the level of regularity.

Presenting the target associations. The target associations were given under the same cover story as in Experiment 1, but with one difference: Each of the plants fed with the candidate substances was fed with two substances at the same time (i.e., the gardener's friend fed each pot of plants two types of substances simultaneously in his yard in a foreign country). In three practice trials, all subjects were presented with the same set of blooming rates involving three known items, rates that were consistent with the knowledge base for both groups.

Subjects were then shown the association between blooming and each of the three pairs of candidate substances: as in Experiment 1, four out of five plants fed with each of the pairs bloomed; in contrast, few of the unfed plants bloomed. For the same reasons as in Experiment 1, except for the abstract causal knowledge manipulated during the learning phase, there should be no difference between groups in performance regarding any candidate items.

Causal rating task. The purpose of this task was to assess subjects' causal judgments on each of the candidate items. They were asked: ''Based on your knowledge about the substances, would you recommend that a customer buy each of these types of substance?'' They were told that the customer would like to make his plants bloom. We assume that recommending a substance implies understanding that it causes blooming. Subjects answered the question on a rating scale that ranged from 4 (''certainly would'' recommend the substance) to −4 (''certainly would not'' recommend it), with ''0'' indicating that the subject did ''not know at all'' whether to recommend it. Their answers were given in a fixed order, starting with the three practice items, followed by items *I*, I_p, *N*, N_p, W_p, and *W*. To measure whether subjects noticed that there was a covariation between a candidate-pair and blooming, they were asked an observation question after the causal rating task: Compared to the gardener's plants that were not fed any substances, did a substantially greater proportion of the five plants fed each pair of these substances bloom?

Results

Categorization Task

The present categorization results replicated those reported in Experiment 1: Subjects were able to form categories defined by the causal features according to contrast; moreover, they did so at the maximal-contrast level. The cue validities of the candidate item in each categorization set, which were identical between groups, cannot explain the observed difference in sorting performance between groups.

Table 2 shows the results for *N*, *W* and *I*. As can be seen, a vast majority of subjects in the color-relevant group sorted item *N* by color, whereas a vast majority of those in the shape-relevant group sorted it by shape, χ^2 (1)

TABLE 2
Percentages of Subjects in the Color-Relevant and Shape-Relevant Groups in Experiment 2 Who Grouped Candidate Item (N, W, or I) with the Color-Matched Item (C-M) or the Shape-Matched Item (S-M) in the Categorization Task

	Item grouped with candidate item	
Group	C-M	S-M
	Item N	
Color-relevant (n = 24)	79	21
Shape-relevant (n = 31)	3	97
	Item W	
Color-relevant (n = 24)	88	13
Shape-relevant (n = 31)	10	90
	Item I	
Color-relevant (n = 24)	100	0
Shape-relevant (n = 31)	45	55

$= 32.9$, $p < .001$. (One subject in each group who sorted this item by neither shape nor color was excluded from the statistical analyses.)

Recall that item *W* shared a color with its color-matched item only at the level of warmth, the level with maximal contrast. A vast majority of subjects in the color-relevant group grouped item *W* with the color-matched item, but almost none of those in the shape-relevant group did so, $\chi^2(1) = 33.3$, $p < .001$. Recall that item *I* shared a shape with its shape-matched item only at the level of regularity. None of the subjects in the color-relevant group grouped this item with the shape-matched item, but most of those in the shape-relevant group did so, $\chi^2(1) = 19.1$, $p < .001$. These results support the criterion of maximal contrast.

Observation and Causal Rating Tasks

As in Experiment 1, all subjects answered the observation questions correctly. For the same reason as in Experiment 1, to test the role of consistency, analyses of the causal ratings for each candidate pair were restricted to subjects who categorized a corresponding item according to the causal dimension in the categorization task. The corresponding item was *N* for pair *N*-N_p; *W* for pair *W*-W_p; and *I* for pair *I*-I_p. (One subject in the shape-relevant group did not perform the rating task because he did not reach that task within the time allowed for the experiment.)

TABLE 3

Mean Recommendation Ratings for Items in Pairs W-W_p and I-I_p Given by Subjects in the Color-Relevant and Shape-Relevant Groups of Experiment 2 Who Sorted the Corresponding Item for the Pair According to the Relevant Causal Dimension in the Categorization Task

	Item	
	Consistent	Inconsistent
Condition	Shape-relevant	Color-relevant
W	2.8 ± .20	−1.9 ± .35
I	2.4 ± .30	−2.4 ± .25
Condition	Color-relevant	Shape-relevant
W_p	2.6 ± .24	−1.6 ± .36
I_p	2.9 ± .16	−0.5 ± .48

Note. For the color-relevant group, the number of subjects included in this analysis was 21 for W and W_p and 24 for I and I_p, for the shape-relevant group, the number of subjects was 27 for W and W_p and 17 for I and I_p.

The present results extend support for the coherence hypothesis to a situation in which the effect of a spurious cause can be explained by an alternative cause. As predicted, for pairs *W*-W_p and *I*-I_p, subjects in both groups were likely to recommend a hierarchically consistent candidate item to a customer, but unlikely to recommend a hierarchically inconsistent one (see Table 3). The hierarchically consistent items for each group were rated *positively* by that group, implying that they were accepted as causal. In contrast, the hierarchically inconsistent items for each group were rated *negatively* by that group, implying that they were not accepted as causal. Recall that, contrary to these results, the approximately equal cue validities of items *N*, *W*, and *I* between groups predict that causal ratings for these items should be *positive* for both groups.

Moreover, these results replicated the interaction between candidate causes and prior causal knowledge reported by Bullock (1979), as the candidate items were causal or noncausal depending on the previously inferred causal relations. For example, as can be seen in Table 3, item *W* was rated positively by the shape-relevant group but negatively by the color-relevant group. In contrast, item W_p was rated negatively by the shape-relevant group and positively by the color-relevant group. The same pattern of results was obtained for pair *I*-I_p. The interaction between candidate item and group is highly reliable: for pair *W*-W_p, $F(1, 46) = 149$ and for pair *I*-I_p, $F(1, 39) = 121$; $p < .001$ for both comparisons.

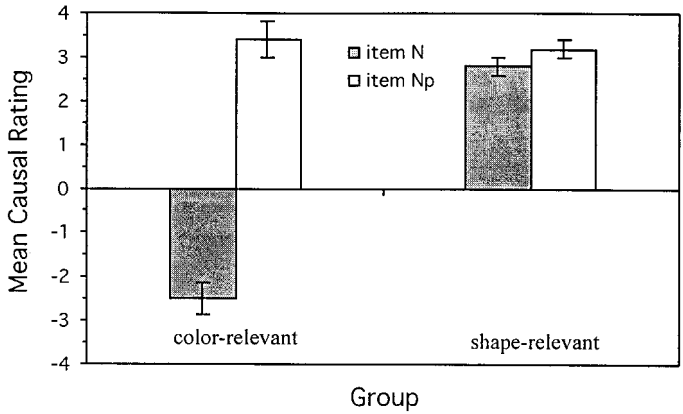


FIG. 9. Mean causal ratings for the novel item N and the known cause N_p given by subjects in the color-relevant and shape-relevant groups in Experiment 2 who sorted item N according to the relevant causal dimension in the categorization task.

Certainty of Causal Judgments

Recall that because a highly plausible candidate is available to explain the effect in this experiment, subjects should have been confident of all their causal judgments, even when the covariation regarding the candidate was inconsistent. Supporting this prediction, the mean causal ratings for all but one of the items were reliably below 0 when the item was hierarchically inconsistent for a group, $t(18) = 6.9$ for *N*, $t(20) = 5.4$ for *W*, $t(23) = 9.6$ for *I*, and $t(26) = 4.4$ for W_p, $p < .001$ for each of these comparisons; and $t(16) = 1.0$ for I_p, n.s. In contrast, the mean causal ratings for all items were reliably above 0 when the item was hierarchically consistent for a group, $t(27) = 14.0$ for *N*, $t(26) = 14.0$ for *W*, $t(16) = 8.0$ for *I*, $t(20) = 10.8$ for W_p, $t(23) = 18.1$ for I_p, $t(18) = 16.2$ for N_p for the color-relevant group, and $t(27) = 16.8$ for N_p for the shape-relevant group, $p < .001$ for each comparison.

Further Evidence Favoring Hierarchical Consistency Over Cue Validity

As explained earlier, under the assumption that subjects judged one and only one item in each pair to be causal, the differing cue validities of W_p and I_p for the two groups could indirectly lead to the observed pattern of ratings for *W* and *I*. But the same argument predicts that *N* and N_p should not both be rated causal.

The causal ratings for the *N*-N_p pair (see Fig. 9) support the coherence hypothesis and contradict the explanation according to cue validity. The known cause N_p , which was hierarchically consistent for both groups, was rated high by both groups (mean ratings of $3.4 \pm .21$ for the color-relevant group and $3.2 \pm .19$ for the shape-relevant group), $t(42) = .4$, $p > .5$. More-

over, the novel item *N*, which was hierarchically inconsistent for the color-relevant group but hierarchically consistent for the shape-relevant group, was unlikely to be recommended by the former group (mean rating of $-2.5 \pm .36$, n = 19), but likely to be recommended by the latter group (mean rating of $2.8 \pm .20$, n = 28), $t(42) = 13.5$, $p < .001$. Note that both *N* and N_p were rated highly causal by the shape-relevant group, contradicting both the cue validity explanation and the prediction regarding the blocking of the learning of *N* according to models that ignore the levels of abstraction of representation.

Discussion

Our coherence hypothesis predicts that in situations in which the effect of a spurious cause can be explained by an alternative cause, people should judge a covariation with relative certainty as either genuinely or spuriously causal depending on its hierarchical consistency. The results of Experiment 2 support this prediction.

At the same time, this experiment showed an interaction between prior causal knowledge and co-occurring candidate causes, replicating the pattern of results in the Jack-in-the-box experiment reported by Bullock (1979; Bullock et al., 1982). Whereas this pattern of results has previously been widely interpreted as contradicting the covariation view, Experiment 2, corroborating Experiment 1, shows that for candidate causes that do not occur independently of alternative causes of an effect, superordinate causal knowledge inferred on the basis of observable frequencies can explain the distinction between genuine and spurious causes.

A Supplementary Test of Alternative Explanations

Recall that for each of the two groups there was a unique item that was presented to only that group. We call these two items the unique item for the color-relevant group (U_c) and for the shape-relevant group (U_s), respectively. These items were associated with different blooming rates: U_s had a high rate, but U_c had a low rate. It might be argued that subjects could simply have made their judgments for the novel items by referring to the respective unique item for each group because they were *a priori* most similar perceptually.

To evaluate this argument, we conducted a spontaneous clustering task analogous to that in Experiment 1 using exactly the same instructions and procedures. In this task, a separate group of subjects from the same pool ($n = 21$) were asked to sort all the stimuli in the present experiment (including the learning and test items for both groups) into groups with respect to either shape or color in turn. If *I* and *W were perceptually most similar to the respective unique items*, subjects in each group should have first grouped these items with the respective unique item. Moreover, because these items

shared values with the respective unique items only at the abstract level of regularity and warmth, these abstract features should have been perceptually salient, implying that subjects would have spontaneously sorted according to these features on the respective dimensions.

The results of the present clustering task show no support for these two predictions. First, the subjects did not have a strong tendency to group U_s with *I* or to group U_c with *W*. When they sorted by *shape*, they took 4.5 trials on average to sort the items into two groups (recall that subjects had to sort all the items into 2 groups by the sixth trial). Four (of 21) never put item *I* and item U_s into one group. Of the remaining subjects, 53% put these items into a group only on their final trial and 82% grouped these two items together no earlier than on their penultimate trial. Only one subject formed a group consisting of only item *I* and item U_s. This pattern of results indicates that to most subjects *I* was not the most similar in shape to U_s, but that instead, *I* was more similar to items in earlier groups to which it belonged. When subjects sorted by *color*, they took 4.6 trials on average to sort the items into two groups. Three (of 21) subjects never put item *W* and U_c in the same group. Of the remaining subjects, 39% put these two items together only on their final trial, and 78% put these items together no earlier than on their penultimate trial. No subject formed a group consisting of only U_c and *W*. These results indicate that item U_c was not perceived to be the most similar to item *W* for any subject. It is therefore very unlikely that the observed differences between groups in the causal ratings for *I* and *W* were due to the spontaneous similarity of these items to the respective unique items for each group.

Second, none of the subjects (out of 20) sorted the items according to regularity when they sorted by shape, and although most of the subjects (16 out of 21) separated blue and green from warmer colors on the 6th trial when they sorted by color, only two of them did so earlier than their fourth trial. These results indicate that the level of abstraction at which *I* and *W* shared values with the respective unique items for each group was not spontaneously salient. Instead, confirming the earlier clustering results, they suggest that these levels, which are predicted by the criterion of maximal contrast, were induced.

GENERAL DISCUSSION

Even when (1) there is no innate causal knowledge about a regularity and (2) conditional contrasts cannot be computed, people are still able to systematically distinguish between genuine and spurious causes. This ability challenges both dominant views of causality: the covariation view and the power view. Whereas the covariation view does not apply, the power view does not explain the nature and origin of the knowledge of causal power

that it assumes, thus pushing the question one level back without answering it. Moreover, neither view can explain the difference in confidence with which people judge a regularity to be spurious depending on whether an alternative cause explains the effect.

We propose an integration of these two views: We assume that superordinate causal knowledge about the candidate cause is critical, but that this knowledge can be induced from observable information. Extending the covariation view, we propose that the level of abstraction of this relation is that with maximal contrast. Finally, extending both views, we propose that the influence of this superordinate knowledge operates in accord with coherence. Causal rules explain regularity, and a reasoner seeks to explain as much as possible with as few rules as possible. This hypothesis predicts that a covariation will be confidently accepted as causal when it is hierarchically consistent, regardless of whether an alternative cause co-occurring with the candidate explains the effect; but, a covariation will be judged as spurious when it is hierarchically inconsistent, with confidence in the judgment being higher when a co-occurring alternative cause explains the effect than otherwise. Results from our two experiments support these predictions.

Relation to Other Previous Work

In addition to the previous work we cited under the covariation view and the power view, three other lines of the work seem particularly relevant to our work.

Perception of Covariation

Cognitive and social psychologists have over and again found that people go beyond the currently given information and are influenced by their knowledge about the world in the perception, interpretation, and comprehension of everyday experience. In their review regarding the perception of covariation, Alloy and Tabachnik (1984) argued that two sources of information, and the consistency between them, determine the perception of covariation. These two sources of information are (1) the situational information about the covariation between the events observed in the current environment and (2) the learner's prior expectations or beliefs about the covariation between the events in question. Our coherence hypothesis extends this line of work in four ways. First, it argues that the level of abstraction of the representation of the prior expectations is critical to the very definition of consistency. Second, it provides a criterion for defining that level. Third, it suggests that prior causal knowledge influences not only the perception of covariation, but also whether the perceived covariation is interpreted as causal. Finally, it argues that coherence as we discussed is important in such interpretations.

Models of Categorization

Judging whether a candidate factor is causal may be regarded as a categorization task: causes would be in one category and noncauses would be in the

complementary category. Various connectionist models have been applied to explain causal inference as well as categorization (e.g., Gluck & Bower, 1988; Krushke, 1992; Rescorla & Wagner, 1972). These models, like virtually all other current connectionist models, eliminate variables and other ''symbolic'' mechanisms (e.g., Elman, 1990; McClelland, Rumelhart, & The PDP Research Group, 1986). But, a fundamental characteristic of causal inference is that reasoners represent causal power as a variable, distinctly from covariation, another variable (Cheng, 1997; Wu & Cheng, 1999). The framework that we adopt suggests that to model causal inference, connectionist models would require an architecture radically different from that of the eliminative variants (see Hummel & Holyoak, 1997, for an example of a *symbolic* connectionist model).

Hierarchical Knowledge and Induction

Reichenbach (1938) proposed that causal laws are nothing but a special case of inductive law, and that induction at a given level of abstraction makes use of information at a superordinate level, a proposal highly similar to ours. He noted that induction involves probabilities at different levels, illustrating his point with the following example. At one time chemists had found that all known substances, except for carbon, melted if they were heated above a certain temperature. Chemists did not believe, however, that carbon was infusible. Rather, they were convinced that the then current technical means simply could not attain the sufficiently high temperature at which carbon would melt. Reichenbach pointed out that in this case, chemists considered inductive inference at two different levels, one of which overrode the other. At the lower level, based on the observation of whether a substance melted as the temperature was gradually increased, one might infer that carbon was infusible but other substances would always be in a liquid state above a certain temperature. However, an inference at a higher level intervened here. Based on the fact that for all the other cases the increase in temperature led to melting, chemists inferred that the same trend would hold for carbon, if the temperature could be further increased. According to Reichenbach, an induction of the higher level always supersedes an induction of the lower one.

The coherence hypothesis generalizes Reichenbach's (1956) proposal to probabilistic events. Moreover, this hypothesis adds an explanation of why confidence accompanying judgments regarding spurious causes using superorindate knowledge should depend on whether a regularity at the lower level is explained by an alternative cause.

What Counts as an Alternative Cause?

We now return to the issue of what counts as an alternative cause when levels of abstraction of the cause are considered. Generally (not just for such considerations), for the purpose of computing conditional contrasts, causes along the *same* causal path or paths to effect *e* as candidate cause *c* should

not count as an alternative cause. If ancestors (i.e., direct or remote causes) of *c* are held constant, ΔP_c with respect to *e* cannot be computed due to lack of variation in *c*. When descendents (i.e., direct or remote effects) of *c* on the causal path(s) to *e* are held constant, ΔP_c with respect to *e*—which becomes 0—does not reveal its (remote) causal power to produce *e* (see the Markov condition in Pearl, 1988, or Spirites et al., 1993). To our knowledge, no psychological research has been conducted to test whether untutored reasoners intuitively conditionalize contrast for a candidate cause only on causes that are not along the same causal paths to an effect.

When levels of abstraction of the candidate cause are considered, when does a factor that is not the candidate cause count as an alternative variable (i.e., an alternative cause, in which case it should be controlled) or an alternative *value* of the same variable as the candidate cause (in which case it is classified as the absence of the candidate cause for the purpose of computing contrast)? For example, when causes of lung cancer are considered for the candidate cause Virginia Slims, should Camels be considered an alternative cause or an alternative value of the same variable, cigarettes?

An assumption in the computation of contrast is that reasoners form variables so that the probability of the effect given one value of a variable can be contrasted with that given the complementary value, while controlling for alternative variables. In our experiments, we constructed materials for which this assumption indisputably holds: blue and green, for example, are psychologically values along the same variable of color; thus, the contrast for ''blue'' classifies ''green'' as nonblue, while controlling for shape. The maximal-contrast criterion is dependent on the distinction (whether innate or learned) between values of the same variable (e.g., blue and green) and values of alternative variables (blue and regularly shaped). Do reasoners make this distinction? If green and other cool colors in our experimental materials, for example, were treated as causes alternative to blue, then the conditional contrast for blue would be the same as that for cool colors, the more abstract level at which many subjects were able to generalize. Our conjecture is that reasoners treat a set of causal factors (1) as values of a single causal variable if they can find a common feature among them that yields a higher contrast with respect to the effect than any of those factors treated as a value, but (2) as alternative causal variables otherwise. Do they? It would be consistent with the coherence hypothesis and the maximal-contrast criterion if they do. A restriction is that the common feature not be ''producing the effect'' because this feature would require first knowing about the effect and would therefore fail to serve the essential purpose of predicting the effect. In our materials, values along the same variable are mutually exclusive for any given entity, whereas values on alternative variables are not.[4] For example, an item cannot be both cool and warm colored,

[4] We thank Clark Glymour for suggesting the mutual exclusivity of values as a criterion for forming a variable.

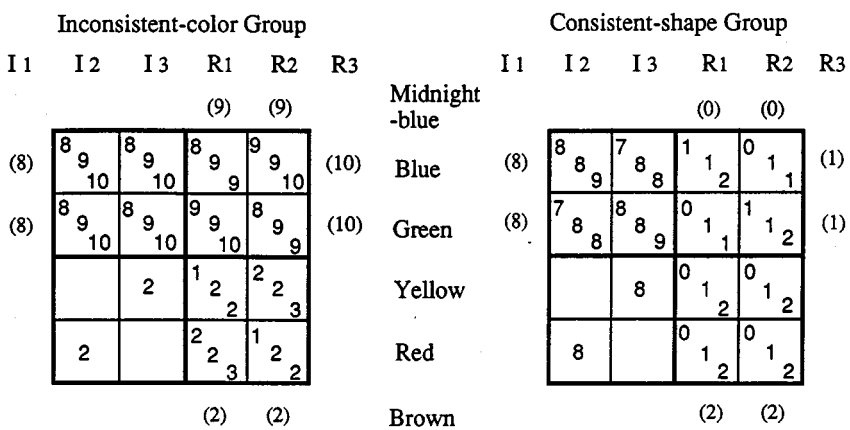
but it can be both cool-colored and regularly shaped. What role does mutual exclusivity play in the formation of variables? To our knowledge, no psychological research has studied these related questions.

CONCLUSION

In our approach to solving the challenging intellectual jigsaw of how people distinguish genuine from spurious causes, we have turned away from the popular debate between the power and covariation views toward an integration and extension of those views. We have noted some gaps overlooked by both previous views and proposed possible pieces to fill some of them. It is clear that we have by no means completed the jigsaw. For example, our proposal regarding maximal contrast has barely been tested, and we have not extended our approach to domains involving continuous variables or geometric information. More generally, coherence clearly has a broader scope within causal inference than what we have discussed here (e.g., see Cheng, 1993). Our aim is merely to fit a new piece, however small, into the puzzle, with the hope that the addition will change the expected shapes of those pieces still missing.

APPENDIX A

The Frequency of Blooming Given as the Knowledge Base of Experiment 1



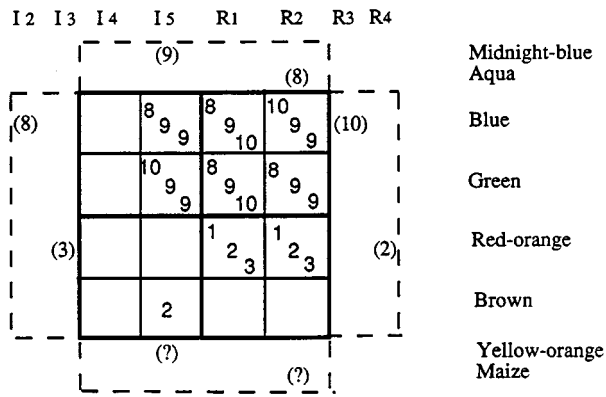
1. Each number in a cell indicates the number of plants (out of 10) fed with each type of substance that bloomed given in the passive learning phase.
2. Each number within parentheses is the number of plants (out of 10) fed with each type of substance that bloomed given as feedback for the dynamic learning phase.

3. The unfed plants bloomed at the rate of 2 in 10 for the inconsistent-color group and 1 in 10 for the consistent-shape group.

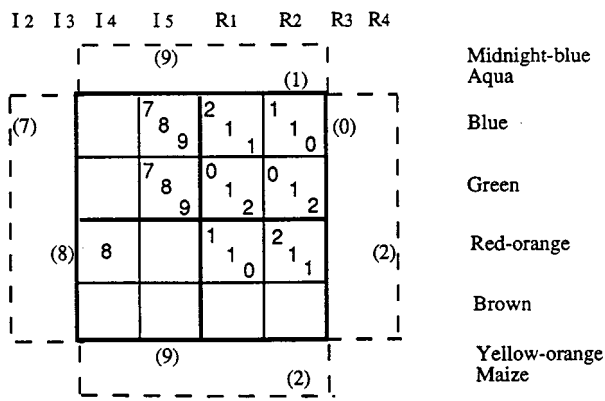
APPENDIX B

The Frequency of Blooming Given as the Knowledge Base of Experiment 2

Color-relevant condition



Shape-relevant condition



1. Each number in a cell indicates the number of plants (out of 10) fed with each type of substance that bloomed given in the passive learning phase.
2. Each number within parentheses is the number of plants (out of 10) fed with each type of substance that bloomed given as feedback in the dynamic learning phase.

3. A question mark in parentheses indicates that no feedback for that type of substance was given.

4. The unfed plants bloomed at the rate of 2 in 10 for the color-relevant group and 1 in 10 for the shape-relevant group.

5. For two ''prediction'' items with novel warm colors but old shapes, feedback was withheld from the color-relevant group (as denoted by the question marks) because these items (which were respectively yellowish orange and maize) were a priori similar in color to candidate item *W* (which was yellow). Feedback on these items might therefore explain performance on *W* for the color-relevant group. But because this difference in feedback occurred after the categorization task, it cannot explain any observed difference in categorization performance. Even for the causal rating task, this difference cannot explain any difference in the inferred knowledge base: withholding feedback regarding these items from the color-relevant group could not possibly have *helped* this group infer that color was causal, and giving feedback regarding these items to the shape-relevant group could not have conveyed any new information to that group—these items had old shapes regarding which subjects were already given an identical blooming rate. Also note that for the shape relevant group, one of these items had a high blooming rate and the other had a low rate. Their blooming rates therefore could not have biased this group's causal ratings for the candidate items one way or the other.

REFERENCES

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

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review,* **91,** 112–149.

Bullock, M. (1979). *Aspects of the young child's theory of causation.* Unpublished doctoral dissertation, University of Pennsylvania.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time.* New York: Academic Press.

Cartwright, N. (1983). *How the laws of physics lie.* Oxford, UK: Clarendon.

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, UK: Clarendon.

Cheng, P. W. (1993). Separating causal laws from casual facts: Pressing the limits of statistical relevance. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 215–264).

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

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

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

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review,* **99,** 365–382.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science,* **14,** 179–211.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science,* **4,** 123–124.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science,* **7,** 155–170.

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

Gluck, M., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General,* **117,** 227–247.

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

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review,* **94,** 211–228.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development,* **60,** 1316–1327.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review,* **99,** 22–44.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition,* **25,** 265–288.

Lewis, C. I. (1929). *Mind and the world order.* New York: Scribner. Chapter XI.

Lien, Y., & Cheng, P. W. (1989). A framework for psychological causal induction: Integrating the power and covariation views. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society,* (pp. 729–733). Hillsdale, NJ: Erlbaum.

Marr, D. (1982). *Vision.* New York: Freeman.

McClelland, J. L., Rumelhart, D. E., & The PDP Research Group. (1986). *Parallel distributed processing; Explorations in the microstructure of cognition: Psychological and biological models* (Vol. 2). Cambridge, MA: MIT Press.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review,* **101,** 587–607.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Hempel, C. (1965). *Aspects of scientific explanation.* New York: Free Press.

Holyoak, K. J., & Hummel, J. E. (in press). The proper treatment of symbols in a connectionist architecture. In E. Deitrich and A. Markman (Eds.), *Cognitive dynamics: Conceptual change in human and machines.* Mahwah, NJ: Erlbaum.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review,* **104,** 427–466

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology,* **3,** 382–407.

Reichenbach, H. (1956). *The direction of time.* Berkeley/Los Angeles: Univ. of California Press.

Reichenbach, H. (1938). *Experience and prediction: An analysis of the foundations and the structure of knowledge.* Chicago, IL: Univ. of Chicago Press.

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* **7**, 573–605.

Salmon, W. C. (1971). Statistical explanation. In W. C. Salmon, R. C., Jeffrey, and J. G. Greeno (Eds.), *Statistical explanation and statistical relevance* (pp. 29–87). Pittsburgh, PA: Univ. of Pittsburgh Press.

Salmon, W. C. (1977). A third dogma of empiricism. In R. Butts & J. Hintikka (Eds.), *Basic problems in methodology and linguistics* (pp. 149–166). Dordrecht: D. Reidel.

Salmon, W. C. (1980). Probabilistic causality. *Pacific Philosophical Quarterly,* **61,** 50–74.

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

Shanks, D., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21). New York: Academic Press.

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

Shultz, T. R., & Kesterbaum, N. R. (1985). Causal reasoning in children. *Annuals of Child Development,* **2,** 195–249.

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

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

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

White, P. A. (1992). Causal powers, causal questions, and the place of regularity information in causal attribution. *British Journal of Psychology,* **83,** 161–188.

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

Wu, M. & Cheng, P. W. (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.