Memory & Cognition
1997, 25 (1), 125–134

Determining whether causal order affects cue selection in human contingency learning: Comments on Shanks and Lopez (1996)

MICHAEL R. WALDMANN
Max Planck Institute for Psychological Research, Munich, Germany

and

KEITH J. HOLYOAK
University of California, Los Angeles, California

Shanks and Lopez (1996) reported three experiments in which they attempted to test whether causal order affects cue selection, and concluded that it does not. Their study provides an opportunity to highlight some basic methodological criteria that must be met in order to test whether and how causal order influences learning. In particular, it is necessary to (1) ensure that participants consistently interpret the learning situation in terms of directed cause–effect relations; (2) measure the causal knowledge they acquire; (3) manipulate causal order; and (4) control the statistical relations between cause and effect. With respect to these criteria, each experiment reported by Shanks and Lopez fails on multiple counts. Moreover, several aspects of the results reported by Shanks and Lopez are explained by causal-model theory, but not by associative accounts. Their study thus adds to a growing body of evidence from different laboratories indicating that human contingency learning can be guided by causal interpretation.

The capacity to acquire causal knowledge is one of our most important cognitive competencies. Many types of animals are clearly capable of learning how to react appropriately to causal contingencies in their environments, thereby satisfying their basic survival goals. A few types of primates, and in particular humans, demonstrate much more sophisticated forms of causal learning and understanding. People do not simply learn to react to causal contingencies; rather, they are capable of imagining the potential effects of causes that are not currently present. In addition to using perceived or imagined causes to predict future effects, people can use perceived or imagined effects as cues to diagnose their unseen causes. These varieties of causal learning and reasoning appear to be central to human abilities to plan actions that will achieve their goals.

The Direction of the Causal Arrow

The centrality of causality in human cognition is reflected in the interest the topic has attracted among both philosophers and psychologists over many centuries. Despite the fact that little agreement has been achieved on how best to conceptualize the fundamental nature of the relationship between causes and effects, one core assumption has been shared by competing theoretical camps: The causal arrow is directed from causes to effects. All psychological theories of causality agree that people conceptualize causes as prior to their effects. Even when people are unable to observe a causal factor, or to perceive a temporal gap between cause and effect, they nonetheless believe that the cause precedes its effect. It is particularly notable that the philosopher David Hume (the forefather of modern associationism) explicitly included the temporal precedence of causes to their effects as part of his definition of causality (see Hume, 1739/1978, p. 173). He recognized a strong conceptual distinction, based in part on temporal asymmetry, between events that are causes and events that are effects.

Modern computer scientists and philosophers have elaborated the central role causal directionality plays in the determination of probabilistic relations among networks of events (Pearl, 1988; Reichenbach, 1956). In addition, cognitive psychologists have obtained evidence that people preferentially code causal contingencies in the cause-to-effect direction, rather than the reverse (Eddy, 1982; Tversky & Kahneman, 1980; Waldmann & Holyoak, 1992). Figure 1 illustrates some consequences of causal directionality for causal inference (Waldmann, Holyoak, & Fratianne, 1995). Figure 1A depicts a *common-cause* structure, in which a single cause produces multiple effects, whereas Figure 1B depicts the complementary *common-effect* structure, in which a number of causes act independently to produce a single effect. A key point to note is that a common-cause structure implies a spurious correlation among the effects. Even though the effects do not directly influence one another, their values will vary

Preparation of this paper was supported by NSF Grant SBR-9310614 to K. Holyoak. We thank T. Wickens for statistical advice. Correspondence should be addressed to M. R. Waldmann, Max Planck Institute for Psychological Research, Leopoldstrasse 24, 80802 Munich, Germany (e-mail: waldmann@mpipf-muenchen.mpg.de).

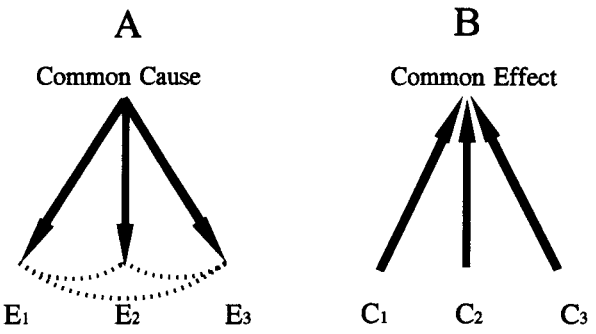

Figure 1. Common-cause structure (A) versus common-effect structure (B). Only the common-cause structure formally implies a spurious correlation (dotted curves) among effects. From "Causal Models and the Acquisition of Category Structure," by M. R. Waldmann, K. J. Holyoak, and A. Fratianne, 1995, *Journal of Experimental Psychology: General, 184*, p. 124. Copyright 1995 by the American Psychological Association. Reprinted with permission.

together as the status of the common cause varies. A classic example is the case of a group of people who become sick after eating together (Reichenbach, 1956). It is natural to explain these co-occurring effects as the consequence of food poisoning (the common cause), rather than as a collection of coincidental illnesses causally unrelated to one another.

In contrast, a common-effect structure (Figure 1B) does not imply a corresponding correlation among the multiple causes. Rather, each cause may vary independently, and their effects may simply summate to determine the value of the effect. For example, several factors (e.g., lack of exercise, fatty diet, and smoking) may separately increase the probability of heart disease. It is possible that multiple causal factors may interact (e.g., perhaps fatty diet and smoking have an "overadditive" impact on heart disease), but in such cases the underlying causal model would have to be elaborated with conjunctive nodes to explicitly code the interactive relations between the causes and the effect. The need for explicit configural features would be expected to increase the difficulty of learning (Dawes, 1988). Thus in common-effect models, sensitivity to correlated causes will require explicit representations of interactive features, whereas in a common-cause model, such sensitivity can emerge implicitly from a causal network based solely on links from individual causes to individual effects (Waldmann et al., 1995).

Waldmann and Holyoak (1992) proposed a causal-model theory, according to which people's acquisition of causal contingencies will be constrained by the causal network that learners impose on the observations. Waldmann and Holyoak observed that people can learn causal connections not only in a *predictive* learning context, where the cues are interpreted as possible causes and the response is the predicted effect, but also in a *diagnostic* learning context, in which the cues are interpreted as effects and the response refers to a diagnosed cause. On the basis of evidence such as that of Tversky and Kahneman (1980), causal-model theory postulates that people will code causal contingencies in the cause-to-effect direction, even when learning takes place in a diagnostic context. One prediction that follows from causal-model theory is that "cue selection" (or competition) will interact with the perceived causal direction of the links. Suppose, for example, that people learn a structure they interpret in terms of Figure 1B, where the cues are the independent causes and the response is the value of the effect. To the extent that one cue is established as a cause of the effect, people will be less likely to attribute causality to other cues that are redundantly paired with this cue during the learning phase. Such cue competition is often termed "blocking" (from the literature on animal conditioning) or "discounting" (from the literature on social attribution), and can be derived both from contingency-based models (see, e.g., Cheng & Novick, 1992; Waldmann & Holyoak, 1992) and the Rescorla–Wagner (1972) model of associative learning. Thus, if an allergic reaction is observed after people eat both strawberries and peanuts, and there is independent evidence that peanuts cause the reaction, observers will be less likely to believe that strawberries cause the reaction (Wasserman, 1990).

On the other hand, suppose that people learn a structure they interpret in terms of Figure 1A, where the cues are the effects and the response is the value of the cause. Waldmann and Holyoak (1992) argued that cue competition among effects need not occur, since multiple independent effects of a common cause do not interact. For example, learning that eating peanuts causes stomach upset would not be expected to block learning that it also causes a skin rash. Waldmann and Holyoak (1992, Experiments 1 and 3) obtained the predicted interaction between perceived causal direction and cue competition (see also Van Hamme, Kao, & Wasserman, 1993). Associative models such as the Rescorla–Wagner model are unable to account for such an interaction, as the cue–response relationship is equated across the variation in perceived causal direction.

What Associationism Forgot

Hume's analysis of causal relations is the intellectual progenitor of all current covariation-based models of causal induction, including causal-model theory (Waldmann & Holyoak, 1992; Waldmann et al., 1995), the probabilistic contrast model (Cheng & Novick, 1992), and associationist accounts based on extensions of the Rescorla–Wagner (1972) model (Gluck & Bower, 1988; Shanks, 1991). However, modern associationist psychology somehow dropped Hume's insight about causal directionality when translating causes and effects into the language of stimuli and responses (see Waldmann, 1996, for more details). Unlike Hume, modern psychological theories of associative learning (e.g., the Rescorla–Wagner theory) describe learning as the acquisition of associative links between *cues* and *outcomes* rather than causes and effects. The organism is conceived of as responding

to cues regardless of what type of events these stimuli actually represent. Shanks and Lopez (1996) present the most recent example of this reductionist program. These investigators claim that people are insensitive to causal directionality, and that they do not differentiate between causes and effects or between predictive and diagnostic inferences. Rather, people simply learn to associate cues with responses, without any sensitivity to whether the cues are understood to be causes, effects, or arbitrary signals devoid of any causal interpretation.

This claim certainly deserves close scrutiny, as it predicts misrepresentations of physical reality that should have detrimental consequences in our daily life. One of the key differences between causes and effects is that causes can be manipulated to achieve effects but effects cannot be manipulated to achieve causes (von Wright, 1971). Failures to distinguish between these two types of events may thus lead to poor judgments in the selection of actions to achieve goals. As pointed out by Reichenbach (1956), the asymmetries between common-cause and common-effect structures are an *empirical* characteristic of the physical world. Accordingly, any theory that posits insensivity to causal directionality implies that people are unable to correctly represent basic features of the physical world they inhabit.

It is important to note that the Rescorla–Wagner learning rule implies an asymmetry between cues and outcomes reminiscent of Reichenbach's (1956) analyses of common-cause and common-effect structures. As pointed out by Van Hamme et al. (1993), the Rescorla–Wagner rule implies competition among cues but not among outcomes. Thus, if the learning situation presents causes as cues and effects as outcomes, the Rescorla–Wagner rule makes the correct predictions regarding asymmetry in competition for cues versus outcomes. Van Hamme et al. (1993) reported experiments in which causes and effects were presented simultaneously so that the causes could be mapped to the cue level and the effects to the outcome level. In this type of learning situation, the Rescorla–Wagner rule indeed predicts that the causes but not the effects should compete, a finding that has been obtained by a number of investigators (Baker & Mazmanian, 1989; Matute, Arcediano, & Miller, 1996, Experiments 1 and 2; Rescorla, 1991; Van Hamme et al., 1993).

However, the Rescorla–Wagner rule yields erroneous predictions for diagnostic learning tasks in which the cues represent effects and the outcomes represent causes. In this kind of situation, the Rescorla–Wagner rule predicts competition among effects but not among causes, a pattern contrary to physical reality. The apparent success of the Rescorla–Wagner rule with predictive learning tasks is therefore dependent on the fortuitous form of the learning task, with causes as cues and effects as outcomes, rather than on the potential of the Rescorla–Wagner rule to adequately represent causal relations. Shanks and Lopez (1996) are clearly aware of the implications of the Rescorla–Wagner rule for diagnostic learning tasks. These investigators study tasks in which effects are presented prior to the causes, and they consequently predict that cue competition should be observed among effects.

METHODOLOGICAL REQUIREMENTS FOR INVESTIGATING CAUSAL DIRECTIONALITY

Associative theories view causal learning as a special case of general contingency learning. According to this class of theories, contingency learning involves the acquisition of associative weights between cues and outcomes regardless of the semantic interpretation of these events. By contrast, causal-model theory is based on analyses of the structural characteristics of *causal* situations, such as patterns of causal directionality. This theory claims that humans are sensitive to these characteristics when learning about causal situations. It certainly would make little sense to claim sensitivity to causal directionality in noncausal learning tasks in which, for example, shapes have to be associated with arbitrary category labels (see, e.g., Shepard, Hovland, & Jenkins, 1961). Thus, causal-model theory focuses on an important *subset* of learning tasks that associative theories try to model. A methodological consequence of this subset–superset relation is that causal-model theory need not prove that it applies to noncausal learning tasks, whereas associative theories are obliged to show that their general theory also applies to the special case of causal learning tasks. Demonstrations that associative accounts successfully model noncausal learning (a claim that is itself questionable with respect to the Rescorla–Wagner model; see Miller, Barnet, & Grahame, 1995) will simply not do. It is therefore important to ensure that the learning tasks indeed involve the acquisition of causal knowledge.

In order to even address the question of whether causal directionality can guide the induction process, it is necessary to design experiments that satisfy a few key methodological requirements. We focus here on four such requirements. In our view, every experiment reported by Shanks and Lopez (1996) fails on multiple counts. In addition, several of the criticisms these investigators directed at the design of the study of Waldmann and Holyoak (1992), which yielded robust effects of causal directionality on cue selection, appear to reflect lack of awareness of these core methodological constraints.

Requirement 1: Ensure That Participants Consistently Interpret the Learning Situation in Terms of Directed Cause–Effect Relations

In order to test whether people's understanding of causal directionality influences their learning, the instructions to participants must lead them to impose a consistent cause–effect interpretation on the materials. If the instructions in an experiment do not make cause–effect relations clear to participants, the results can shed no light on the question of whether causal interpretation can guide learning.

To meet this requirement, Waldmann and Holyoak (1992) used cover stories that clearly conveyed a causal asymmetry: for example, a virus (cause) that affects people's appearance (effect), or an alarm switch (cause) that turns on an alarm (effect). In contrast, Shanks and Lopez (1996, Experiments 1 and 3; following Shanks, 1991) employed materials based on artificial "diseases" and "symptoms," simply assuming that diseases must be causes and symptoms must be effects. But as Waldmann and Holyoak (1992, note 1; see also Melz, Cheng, Holyoak, & Waldmann, 1993) have pointed out, calling cues "symptoms" is not sufficient to unambiguously specify the underlying causal model so as to allow clear predictions to be derived. In actual medical reasoning, observable symptoms or signs may be interpreted as causes rather than effects (e.g., puncture wounds may be the cause of blood poisoning, rather than the reverse). Symptoms may also be links in a causal chain, thus serving as both causes and effects (e.g., fever may be the effect of an infection and the cause of dehydration). Moreover, a "disease" is not unambiguously a cause; rather, it may simply name a collection of symptoms that constitutes a syndrome. For example, "AIDS" is the name applied to a collection of symptoms (e.g., weight loss, susceptibility to infection) believed to be caused by the HIV virus. It is the virus, not the disease label, that is considered to be the causal factor.

The cover stories employed by Shanks and Lopez (1996, Appendix) to instantiate a diagnostic learning condition (in which cues are effects and the response is a cause) failed to clarify the causal relations to participants. In their concrete EC condition (where "EC" signifies that effects were presented as cues and causes as outcomes), the cues were symptoms such as "slurred speech" and the responses were artificial diseases such as "Phipp's syndrome." The instructions did not indicate whether Phipp's syndrome, for example, was to be interpreted as a cause, rather than simply a name for a collection of symptoms (although the term "syndrome" suggests the latter interpretation). The fact that the participants in their EC conditions stated in interviews that "in the real world... the cues would be effects and the outcomes the causes" (p. 517) does not necessarily imply that the cues were interpreted as *independent* effects of the novel invented diseases presented in the experiments, as the causal status of a symptom is certainly not invariant across different diseases. The cover stories for the abstract EC condition were even less clear, consisting of the following: "You will be shown the symptoms [abstract group only: which are labeled by the letters A to N] that each patient has, and then asked to say which illness you think the person is suffering from. [Abstract group only: Some of these people have Disease 1, some Disease 2...]" (p. 522). This minimalist context provides essentially no information about causal relations, making it very likely that participants simply treated the task as involving the learning of arbitrary contingencies between letters and numbers. Clearly, if the experimenters' instructions fail to provide any information about cause–effect relations, participants will have no basis for learning contingencies in the cause-to-effect direction. It is likely that in learning tasks in which no clear causal interpretation is provided, participants may resort to the default assumption that the cues presented first represent causes, as this interpretation would correspond to the natural order of events in the real world. Melz et al. (1993) have shown that Shanks's (1991) experiments can be modeled by a contingency theory when the symptoms are coded as causal factors. The instructions and materials in those experiments are similar to the ones used by Shanks and Lopez (1996). However, unlike Shanks and Lopez (1996), Shanks (1991) never claimed that the symptoms presented in the learning tasks represent effects (see also Shanks, 1993).

Not surprisingly, given that the abstract conditions used by Shanks and Lopez (1996) in their Experiment 1 were essentially devoid of causal content, their results yielded no effect of causal directionality for these conditions. Inspection of the results (see their Figure 1) suggests, however, that causal directionality did have an influence in the concrete conditions, for which some participants may have interpreted the EC condition as involving effects as cues and causes as responses, despite the ambiguous cover stories. The size of the cue competition effect (i.e., the difference between the mean rating for contingent vs. noncontingent[1] cues) was substantially larger in the (predictive) concrete CE condition than in the (diagnostic) concrete EC condition (41 vs. 14). Using the mean square error reported by Shanks and Lopez (608.3), we conducted a test of the simple interaction effect for the concrete condition; this test yielded a significant difference [$t(60) = 2.19, p < .05$].[2] Thus despite the deficiencies of their causal cover stories, Shanks and Lopez (1996) in fact obtained the type of interaction that is predicted by causal-model theory and that cannot be accounted for by the Rescorla–Wagner model.

In their Experiment 2, Shanks and Lopez (1996) ran the abstract EC condition, but with a more credible causal cover story. However, in Experiment 3, they reverted to the weaker cover story used in Experiment 1.

Shanks and Lopez (1996) not only failed to provide clear causal cover stories in the instructions for their own experiments, but they also criticized Waldmann and Holyoak (1992) for having done so. According to Shanks and Lopez, "the interaction Waldmann and Holyoak (1992, Experiment 3) obtained arose in an experiment in which rather different cover stories were given to subjects in the CE and EC tasks" (p. 514). Indeed, Waldmann and Holyoak derived predictions for common-cause versus common-effect structures and clearly provided instructions suggesting these two types of structures. In Waldmann and Holyoak's (1992) Experiment 3, the participants' task was to learn to predict the state of an alarm on the basis of the state of buttons. These buttons were either defined as causes of the alarm (predictive learning) or as effects of the alarm (diagnostic learning). The participants received identical learning information

in both conditions so that they observed a perfect correlation between button P and the state of the alarm in Phase 1, and a perfect correlation of the buttons P, R, and the state of the alarm in Phase 2. In both learning conditions, these two buttons were characterized as being located in different rooms, and as being either pressed (predictive learning) or observed (diagnostic learning) by two different people from these rooms. Thus in the predictive conditions, the states of the buttons were described as potential causes of a common effect (Figure 1B), whereas in the diagnostic condition they were described as potential effects of a common cause (Figure 1A).

Shanks and Lopez (1996) concluded that in Waldmann and Holyoak's (1992) predictive condition "there is no reason why subjects should have believed that there was any necessary relationship between pressings of Buttons P and R" (p. 514).[3] This is precisely the analysis offered by causal-model theory, as indicated by the independent causes illustrated in Figure 1B. For the diagnostic condition, on the other hand, Shanks and Lopez suggested that "subjects might have reasoned that Light R was perfectly correlated with Light P via some causal link, in which case equivalent judgments should have been given to them" (p. 514). Again, this is simply a restatement of causal-model theory. As illustrated in Figure 1A, a common cause implicitly generates a correlation among its multiple effects. As pointed out by Reichenbach (1956), multiple effects of a common cause are spuriously correlated, and can be "screened off" by holding the common cause constant, whereas multiple causes of a common effect are not necessarily correlated, and cannot be rendered conditionally independent by their joint effect. Their analysis of Waldmann and Holyoak's (1992) Experiment 3 thus reveals that Shanks and Lopez (1996) are themselves sensitive to the structural implications of the causal arrow, even though they prefer not to attribute such sensitivity to the participants in their experiments.

In summary, Shanks and Lopez's (1996) Experiments 1 and 3 failed to meet Requirement 1, as the cover stories did not convey clear information about causal relations; nonetheless, the concrete conditions (used only in their Experiment 1) yielded the interaction between cue competition and causal direction predicted by causal-model theory.

Requirement 2: Measure the Causal Knowledge That Participants Acquire

In order to assess whether causal order influences the knowledge of causal relations that people acquire, it is clearly necessary to measure what they learn about such relations. For example, Waldmann and Holyoak (1992, Experiment 1), Matute et al. (1996, Experiments 1–2), and Van Hamme et al. (1993) asked their participants to rate how sure they were that a factor was a cause (or effect); Waldmann and Holyoak (1992, Experiments 2–3) had participants rate the degree to which a factor was "predictive." In contrast, Shanks and Lopez (1996; see also Shanks, 1991) chose to ask, "How strongly is [cue] associated with [outcome]?"

Shanks and Lopez (1996) seem to believe that the latter question is semantically equivalent to the predictiveness questions used by Waldmann and Holyoak (1992). Most saliently, the *y*-axis of their Figure 1, which presents the mean association ratings obtained in Experiment 1, is labeled "Mean Rating of Predictiveness," even though no such ratings were obtained. In fact, Shanks and Lopez never asked the participants in any of their three experiments to assess the predictiveness of cues, but instead asked for ratings of associations.

However, it is implausible that association ratings and predictiveness ratings provide equivalent measures of causal relations. The results of Waldmann and Holyoak (1992) indicate that the type of test question can strongly influence the ratings given by participants. In that study, a very different pattern of ratings was obtained for identical learning materials when people were asked whether the cues were effects (Experiment 1), versus whether the cues were predictive of the outcomes (Experiment 2). Causal-model theory predicts that participants who are asked to give *diagnostic* judgments (e.g., predictiveness ratings) are sensitive to whether an effect is potentially caused by one or by several causes. For example, a symptom such as fever may be a deterministic effect of a disease; nonetheless, its diagnostic value for any particular disease crucially depends on whether there are alternative causes. Since fever is a symptom of many diseases, it is a bad diagnostic sign for any particular one. Accordingly, causal-model theory predicts that participants would tend to give low ratings when asked about how predictive fever is for the new disease. However, it seems likely that participants would nevertheless view fever as being highly *associated* with the flu.[4] In everyday usage, "associations" are bidirectional (e.g., bread is associated with butter and vice versa), and hence are unlikely to reflect diagnostic effect–cause relations that are sensitive to alternative causal factors. To the best of our knowledge, no systematic studies have been conducted that investigate the relation between association ratings and the observed causal relations, but it seems questionable to assume that association ratings provide a direct measure of diagnostic inferences.

Causal-model theory postulates a single learning mechanism by which people acquire cause–effect structures; people appear able to flexibly access such structures in both the cause–effect and the effect–cause direction, depending on the question they are asked. To test the theory, therefore, careful attention must be paid to the nature of the questions and context used to assess participants' causal knowledge. For example, Matute et al. (1996, Experiment 3) present evidence suggesting that participants are also able to give *relative* assessments of causal strengths. These investigators have shown that with some test questions, participants tend to rate a poor diagnostic cue higher in the context of even poorer alternative cues, as opposed to a context that includes more valid cues. An

experiment might thus yield apparent cue competition in a diagnostic condition because people interpret the question as requiring a judgment of relative rather than absolute diagnostic validity. It follows that the set of alternatives being evaluated must also be controlled.

Of course, Shanks and Lopez (1996) might claim that association ratings, cause ratings, predictiveness ratings, and so on, all measure the same underlying variable, which can be modeled by a single associative weight. However, this claim clearly runs counter to the available empirical evidence indicating that rating patterns vary with the semantic form of the test question (see Van Hamme et al., 1993, p. 806, for further discussion of this point). Accordingly, Shanks and Lopez (1996) cannot justify their claim to have obtained evidence contrary to the results of Waldmann and Holyoak (1992) when they did not ask the same question of the participants in their experiments. Labeling a dependent measure on the basis of association ratings "Mean Rating of Predictiveness" is no substitute for asking participants to give ratings of predictiveness.

Requirement 3: Manipulate Causal Order

The most transparent requirement for assessing whether causal order affects cue selection is, of course, to vary causal order. It is therefore remarkable that of the three experiments reported by Shanks and Lopez (1996), only Experiment 1 included causal order as a variable. All demonstrations of the impact of causal order have focused on the interaction between causal order and cue competition (i.e., the reduction in cue competition that accompanies a switch in interpretation of the cue–response from cause–effect to effect–cause; Van Hamme et al., 1993; Waldmann & Holyoak, 1992). In fact, as noted above, Shanks and Lopez (1996) actually replicated such an interaction for the concrete conditions in their Experiment 1. By failing to vary causal order in their later experiments, they precluded the possibility of obtaining additional evidence in favor of the phenomenon they wanted to dismiss.

Shanks and Lopez (1996) instead preferred to focus on the absence of cue competition in the diagnostic condition of Waldmann and Holyoak (1992, Experiment 3), rather than on the observed interaction between the predictive and the diagnostic condition. It is indeed the case that causal-model theory predicts that in a well-executed experiment it should be possible not only to reduce cue competition in a diagnostic common-cause relative to an otherwise identical predictive common-effect learning context, but also to eliminate cue competition in the diagnostic condition. Such an experiment would have to ensure that (1) participants who are required to give diagnostic judgments do not believe there are multiple alternative causes of the diagnostic signs and (2) all participants in each condition consistently interpret the causal relation in a single direction (either cause–effect or effect–cause). Unfortunately, there is no reason to suppose that any of the experiments performed by Shanks and Lopez satisfied either of these conditions. In particular, (2) is equivalent to Requirement 1, which we have argued was not satisfied by their Experiments 1 and 3. In the limit, it is only necessary for a single participant in the diagnostic learning condition to fail to impose an effect–cause direction on the cue–response in order to create a nonzero "cue selection effect" in that condition. Even an experiment with carefully constructed causal cover stories could easily fall victim to occasional misinterpretation of the materials. When the causal cover stories are vague to nonexistent, as in Experiments 1 and 3 of Shanks and Lopez (1996), such misinterpretations by participants are assured.

The only way to provide a fair test of whether causal order affects cue selection, therefore, is the obvious one: Vary causal order and see if it makes a difference. Only the first of the three experiments of Shanks and Lopez (1996) even attempted such a test, and the results for the concrete conditions (despite their other methodological problems) did reveal a difference in cue selection.

Requirement 4: Control the Statistical Relations Between Cause and Effect

As Shanks and Lopez (1996) correctly noted, causal-model theory assumes (following statistical relevance theories; Allan & Jenkins, 1980; Cheng & Novick, 1992; Salmon, 1971) that, in situations with a single cause, participants compute the degree of statistical contingency, or contrast, between cause and effect, defined as the difference

$$\Delta P = P(E|C) - P(E|{\sim}C)$$

—that is, as the difference between the conditional probability of a Target Effect E given the presence of a potential Causal Factor C and its probability given the absence of the factor ($\sim C$). Thus, in both predictive and diagnostic learning contexts, all predictions about cue competition depend on equating the statistical relations between cause and effect.

Shanks and Lopez (1996) criticized the design used by Waldmann and Holyoak (1992) (a "blocking design"[5] modeled after that of Kamin, 1968), and instead used other designs that permit a comparison between a condition which (according to associative theories) should exhibit cue competition with a condition in which no cue competition is predicted. In the contingent condition of Experiments 1 and 2, for example, the participants observed that Cues C and D were constantly paired with the outcome, and that Cue D by itself was paired with the absence of the outcome. In the noncontingent condition, A and B were paired with a different outcome, but B by itself was also paired with this outcome. The crucial comparison involves the Cues A and C, which had been paired with their respective outcomes for an equal number of trials. Associative theories predict cue competition between Cues A and B in the noncontingent condition,

but not between C and D in the contingent condition; hence Cue C should be rated higher than Cue A. Shanks and Lopez (1996) apparently performed these experiments in the belief that causal-model theory predicts that Cue C should be rated as more causal than Cue A in the predictive but not in the diagnostic learning context.

But the above design confounds "cue selection" with contrast, and hence it does not address the question of whether cue selection is affected by causal order. If in the diagnostic effect–cause (EC) condition the cues are interpreted as effects and the outcomes as causes, as Shanks and Lopez (1996) intended their participants to do, the contingency between the cause and the effect for Cue A (noncontingent condition) is 0.5, whereas that for Cue C (contingent condition) is 1.0. Accordingly, causal-model theory predicts a difference in the ratings for Cues A and C in the same direction as is predicted by associative accounts.

Given the confounding between cue competition and contrast in the design used by Shanks and Lopez (1996), why did these investigators obtain an interaction between cue competition and causal order in the concrete conditions of their Experiment 1, as we reported above? A more detailed examination of the relevant contrasts provides some insight. For those participants who in fact interpreted the cue–response as effect–cause in the concrete EC condition, we have seen that the relevant contingencies are 1.0 (contingent condition) and 0.5 (noncontingent condition). As elaborated by Melz et al. (1993), the corresponding predictive (CE) condition involves the computation of *conditional* contingencies in which potential co-factors are held constant. In the contingent condition, the appropriate assessment of the causal status of Cue C requires holding Co-factor D fixed, whereas in the noncontingent condition, Cue A should be evaluated when Co-factor B is held constant (see Shanks & Lopez, 1996, Table 1). The conditional contingency between C and the outcome in the presence of D is 1.0, whereas the conditional contingency between A and the outcome in the presence of B is 0. (See Melz et al., 1993, for more detailed analyses of similar experiments reported by Shanks, 1991.) Thus, causal-model theory predicts that Cue C will be perceived as more causal than Cue A in both the predictive and the diagnostic conditions; however, the difference should be larger in the predictive (CE) direction based on contrasts of 1.0 and 0, respectively, for Cues C and A, than in the diagnostic (EC) direction based on corresponding contrasts of 1.0 and .5. The data for the concrete conditions of Shanks and Lopez (1996, Experiment 1) confirm this prediction, which is not made by associative models.

In their Experiment 3, Shanks and Lopez (1996) changed the design to avoid the confounding. In order to equate contingency, trial types were added in Experiment 3 that (according to associative models) should not affect cue competition between the other cues (see Shanks & Lopez, 1996, Table 3). Only the abstract EC condition was run, and the results revealed a small but statistically significant rating difference between the two critical cues.

This finding, of course, provides no evidence that "causal order does not affect cue selection" (the claim made in the title of the Shanks & Lopez, 1996, article). We have already seen that Experiment 3 failed to meet any of the first three methodological requirements: The abstract EC condition did not establish a clear causal interpretation of the materials; the dependent measure was an association rating rather than a direct causal judgment; and causal order was not even varied.

In fact, causal-model theory can actually explain certain aspects of the results of Experiment 3 that contradict associative accounts. First, as Shanks and Lopez (1996) emphasized, associative models predict that the design changes introduced in Experiment 3 relative to Experiments 1–2 should have had no effect on the ratings for the critical cues. However, a comparison between the results for the abstract EC condition of Experiment 1 versus Experiment 3 (conditions based on identical instructions and materials) reveals an apparent interaction. In Experiment 1, the mean ratings were 78.3 for the contingent condition versus 56.3 for the noncontingent condition, a difference of 22 on a scale from 0 to 100. The corresponding means in Experiment 3 were 66.6 and 58.2, a difference of just 8.4. Of course, this cross-experiment comparison only provides suggestive hints rather than clear proof. But it is instructive to see that this reduction in the size of the "cue selection effect" across the two designs is consistent with a contingency analysis, since the difference between the cause–effect contingencies for the critical cues was 0.5 in Experiment 1 versus 0 in Experiment 3. This cross-experiment interaction suggested by the results of Shanks and Lopez (1996), which is inconsistent with associative models, can be predicted by models such as causal-model theory, which emphasizes sensitivity to cause–effect contingencies.

Second, causal-model theory provides an explanation of the otherwise puzzling difference in ease of learning the noncontingent versus contingent structures used in Experiment 3 of Shanks and Lopez (1996). The mean percentage of correct responses across the final two learning trials was 80.9 in the noncontingent condition versus just 69.1 in the contingent condition. In contrast, the comparable conditions in Experiments 1 and 2 were about equal in difficulty (89.1 vs. 86.0 in Experiment 1; 89.1 vs. 89.5 in Experiment 2).

From the point of view of causal-model theory, this difference in ease of learning reflects a new confounding between cue selection and category structure created by the altered design that Shanks and Lopez (1996) used in their Experiment 3. Causal models generally have structural implications that may or may not be compatible with the observed learning input. Waldmann et al. (1995) presented five experiments demonstrating that learning difficulty is influenced by the fit between the structural

implications of the instructed causal models and the structure of the learning input (see also Waldmann & Holyoak, 1990). A comparison of the two conditions of Shanks and Lopez's (1996) Experiment 3 reveals a serious confounding with category structure.

Described in the cause–effect direction, the noncontingent condition presented the following learning input: Cause 1 → AB, Cause 1 → B, no-cause → C. This structure is compatible with a simple common-cause model of the sort depicted in Figure 1A, in which Cause 1 deterministically produces Symptom B, and weakly produces Symptom A. The contingent condition, however, presented a learning input that is incompatible with any simple common-cause structure: Cause 2 → DE, Cause 2 → F, no-cause → E. This rather unusual structure exhibits a situation in which a single cause has disjunctive effects. The disease (i.e., Cause 2) either causes the Symptom Complex DE, or else it causes the Symptom F, but no other combinations of D, E, and F are ever observed. This learning input is clearly incompatible with a causal model that simply links a common cause to three independent effects, as depicted in Figure 1A. If we assume that people tend to initially apply simple causal models such as a common-cause model, then the initial model would have to be modified to account for the peculiar interaction of the effects (see Waldmann et al., 1995, for detailed analyses of such cases). Hence causal-model theory predicts that the simple causal structure embodied in the noncontingent condition should be easier to learn than the disjunctive structure embodied in the contingent condition, a prediction confirmed by the results of Shanks and Lopez (1996).

In summary, Experiments 1 and 2 of Shanks and Lopez (1996) featured a design that confounds cue selection with statistical contrast. Experiment 3 featured a design that instead confounds the two conditions with a simple versus disjunctive cause–effect relation.

HOW DOES CAUSAL ORDER GUIDE INDUCTION OF CAUSAL RELATIONS?

We have highlighted four basic methodological requirements that must be met in order to address the question of whether causal order affects learning: (1) ensuring that participants consistently interpret the learning situation in terms of directed cause–effect relations; (2) measuring the causal knowledge they acquire; (3) manipulating causal order; and (4) controlling the statistical relations between cause and effect. Each of the three experiments reported by Shanks and Lopez (1996) violated at least three of these four methodological constraints. Nonetheless, their Experiment 1 yielded an unreported interaction between causal order and cue competition for their concrete conditions; in addition, cross-experiment comparisons indicate that other aspects of their results are explained by causal-model theory, but not by associative accounts. Their study thus adds to a growing body of evidence from different laboratories indicating that human contingency learning can be guided by causal interpretation (Cheng, Park, Yarlas, & Holyoak, 1996; Matute et al., 1996; Melz et al., 1993; Van Hamme et al., 1993; Waldmann, 1996; Waldmann & Hagmayer, 1995; Waldmann & Holyoak, 1990, 1992; Waldmann et al., 1995).

Although the role of causal order in guiding learning has been empirically established, its theoretical interpretation remains in dispute. Shanks and Lopez (1996), who adopt the standard Rescorla–Wagner approach of mapping effect cues to the input level in diagnostic learning tasks, suggested that the results of Waldmann and Holyoak (1992, Experiment 1) might be accounted for by assuming that participants acquire and run associative networks in both directions. Predictive inferences (e.g., from disease to symptoms) would then be based on a cause–effect network, whereas diagnostic inferences would be based on an effect–cause network. Waldmann and Holyoak (1992, p. 233) outlined some of the problems faced by such proposals to elaborate associative models to make them consistent with the observed impact of causal order on learning. As no one has yet actually produced a computational version of such complex networks, they remain speculative possibilities. In addition, the dual-network approach does not appear to provide an acount of why diagnostic inferences are sensitive to the availability of alternatives causes of effects (Waldmann, 1996; Waldmann & Holyoak, 1992).

On the basis of the empirical evidence available, it seems that associative theories do not adequately model the subset of learning tasks that require the acquisition of causal knowledge. It might still be the case, of course, that associative theories such as the Rescorla–Wagner theory can adequately model noncausal contingency learning. But in light of recent assessments, it seems unlikely that the Rescorla–Wagner theory can provide a complete account of associative learning (Gallistel, 1990; Miller et al., 1995). Even with apparently noncausal tasks, contingency theories may provide a more promising approach. When the context does not establish any clear causal interpretation, the default assumption of the learner may be that the events presented first actually represent causes. On the basis of this assumption, Cheng and Holyoak (1995; Cheng et al., 1996) have shown that contingency theories predict many findings that present problems for the Rescorla–Wagner theory.

It is safe to assume that the theoretical interpretation of the influence of causal order and other structural aspects of cause–effect relations will remain a focus of theoretical debate for some time. The evidence to date, however, provides compelling reasons to believe that associationism made an error when it forgot Hume's insight about the temporal asymmetry of cause and effect. Our goal in the present paper is to prevent even more basic errors—forgetting to control, measure, and manipulate causal order in an unconfounded fashion when investigating its role in learning.

REFERENCES

ALLAN, L. G., & JENKINS, H. M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology*, **34**, 1-11.

BAKER, A. G., & MAZMANIAN, D. (1989). Selective associations in causality judgments II: A strong relationship may facilitate judgments of a weaker one. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 538-545). Hillsdale, NJ: Erlbaum.

CHAPMAN, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 837-854.

CHAPMAN, G. B., & ROBBINS, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, **18**, 537-545.

CHENG, P. W., & HOLYOAK, K. J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In H. L. Roitblat & J.-A. Meyer (Eds.), *Comparative approaches to cognitive science* (pp. 271-302). Cambridge, MA: MIT Press.

CHENG, P. W., & NOVICK, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, **99**, 365-382.

CHENG, P. W., PARK, J.-Y., YARLAS, A. S., & HOLYOAK, K. J. (1996). A causal-power theory of focal sets. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Causal learning* (Vol. 34, pp. 313-355). San Diego: Academic Press.

DAVEY, G. C. L., & SINGH, J. (1988). The Kamin "blocking" effect and electrodermal conditioning in humans. *Journal of Psychophysiology*, **2**, 17-25.

DAWES, R. M. (1988). *Rational choice in an uncertain world*. San Diego: Harcourt Brace Jovanovich.

EDDY, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). New York: Cambridge University Press.

GALLISTEL, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

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

HUME, D. (1739/1978). *A treatise of human nature*. Oxford: Oxford University Press, Clarendon Press.

LOVIBOND, P. F., SIDDLE, D. A. T., & BOND, N. (1988). Insensitivity to stimulus validity in human Pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, **40B**, 377-410.

MARTIN, I., & LEVEY, A. B. (1991). Blocking observed in human eyelid conditioning. *Quarterly Journal of Experimental Psychology*, **43B**, 233-256.

MATUTE, H., ARCEDIANO, F., & MILLER, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 182-296.

MELZ, E. R., CHENG, P. W., HOLYOAK, K. J., & WALDMANN, M. R. (1993). Cue competition in human categorization: Contingency or the Rescorla–Wagner learning rule? Comments on Shanks (1991). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1398-1410.

MILLER, R. R., BARNET, R. C., & GRAHAME, N. J. (1995). Assessment of the Rescorla–Wagner model. *Psychological Bulletin*, **117**, 363-386.

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

REICHENBACH, H. (1956). *The direction of time*. Berkeley: University of California Press.

RESCORLA, R. A. (1991). Associations of multiple outcomes with an instrumental response. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 465-474.

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. Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

SALMON, W. (1971). Statistical explanation. In W. Salmon (Ed.), *Statistical explanation and statistical relevance* (pp. 29-87). Pittsburgh: University of Pittsburgh Press.

SHANKS, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology*, **37B**, 1-21.

SHANKS, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 433-443.

SHANKS, D. R. (1993). Associative versus contingency accounts of category learning: Reply to Melz, Cheng, Holyoak, and Waldmann (1993). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1411-1423.

SHANKS, D. R., & LOPEZ, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory & Cognition*, **24**, 511-522.

SHEPARD, R. N., HOVLAND, C. L., & JENKINS, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**(13, Whole No. 517).

TVERSKY, A., & KAHNEMAN, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 49-72). Hillsdale, NJ: Erlbaum.

VAN HAMME, L. J., KAO, S.-F., & WASSERMAN, E. A. (1993). Judging interevent relations: From cause to effect and from effect to cause. *Memory & Cognition*, **21**, 802-808.

VON WRIGHT, G. H. (1971). *Explanation and understanding*. Ithaca, NY: Cornell University Press.

WALDMANN, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Causal learning* (Vol. 34, pp. 47-88). San Diego: Academic Press.

WALDMANN, M. R., & HAGMAYER, Y. (1995). Causal paradox: When a cause simultaneously produces and prevents an effect. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 425-430). Mahwah, NJ: Erlbaum.

WALDMANN, M. R., & HOLYOAK, K. J. (1990). Can causal induction be reduced to associative learning? In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 190-197). Hillsdale, NJ: Erlbaum.

WALDMANN, M. R., & HOLYOAK, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, **121**, 222-236.

WALDMANN, M. R., HOLYOAK, K. J., & FRATIANNE, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, **124**, 181-206.

WASSERMAN, E. A. (1990). Attribution of causality to common and distinctive elements of compound stimuli. *Psychological Science*, **1**, 298-302.

NOTES

1. The term *noncontingent* is a misnomer, as the structure for this condition actually exhibited a moderate positive contingency, both in the cue-to-response and response-to-cue direction.

2. As the interaction between causal directionality and cue competition is a specific prediction of causal-model theory, it seems appropriate to treat tests of this effect as a planned comparison for each set of stimulus materials.

3. Shanks and Lopez (1996) asserted that the simultaneous pressing of the buttons (predictive condition) was explicitly described as accidental in the instructions used by Waldmann and Holyoak (1992). This is a misunderstanding. The participants were not informed in advance about the structure of the learning items; rather, the correlation between Buttons P and R was simply observed during learning in both the predictive and the diagnostic conditions.

4. Shanks and Lopez (1996) believe that causal-model theory predicts cue competition in diagnostic learning contexts when the cues are concrete but not when they are abstract. What Waldmann and Holyoak (1992) actually claimed was that people will be sensitive to whether an effect is potentially caused by one or by several causes. This distinction is conceptually orthogonal to the abstractness issue.

A concrete symptom may be very diagnostic of a disease (when there are no alternative causes), and an abstract symptom may be a bad diagnostic sign (when the experimental task presents a causal structure in which this symptom is caused by several diseases) (see Waldmann, 1996).

5. Shanks and Lopez (1996) argued that Waldmann and Holyoak's (1992) choice of a blocked- rather than an intermixed-trials design favored absence of cue selection. To support this claim, they exclusively cited research that failed to obtain reliable blocking effects with low-level learning tasks such as eyelid conditioning (Davey & Singh, 1988; Lovibond, Siddle, & Bond, 1988; Martin & Levey, 1991). It is certainly interesting that cue competition seems to be more reliably observed in high-level learning tasks than in such low-level tasks, suggesting that attempts to reduce causal induction to conditioning mechanisms are indeed misguided. However, Shanks and Lopez appear to have overlooked the fact that Waldmann and Holyoak demonstrated an *interaction* between cue competition and causal order *within* the blocking design: With identical learning input, cue competition was observed in the predictive but not in the diagnostic condition. It should also be noted that (1) cue competition in predictive blocked-trial tasks has been observed in many previous studies of causal induction (Chapman, 1991; Chapman & Robbins, 1990; Shanks, 1985), and (2) the interaction of cue competition with causal order has been observed in experiments that do not use a blocking paradigm (Matute et al., 1996, Experiments 1 and 2; Van Hamme et al., 1993).

(Manuscript received May 31, 1995;
revision accepted for publication July 26, 1995.)