Journal of Experimental Psychology:
Learning, Memory, and Cognition
2000, Vol. 26, No. 1, 53-76

0278-7393/00/$5.00 DOI: 10.1037//0278-7393.26.1.53

Competition Among Causes But Not Effects in Predictive and Diagnostic Learning

Michael R. Waldmann
University of Göttingen

Causal asymmetry is one of the most fundamental features of the physical world: Causes produce effects, but not vice versa. This article is part of a debate between the view that, in principle, people are sensitive to causal directionality during learning (causal-model theory) and the view that learning primarily involves acquiring associations between cues and outcomes irrespective of their causal role (associative theories). Four experiments are presented that use asymmetries of cue competition to discriminate between these views. These experiments show that, contrary to associative accounts, cue competition interacts with causal status and that people are capable of differentiating between predictive and diagnostic inferences. Additional implications of causal-model theory are elaborated and empirically tested against alternative accounts. The results uniformly favor causal-model theory.

Causal prediction tasks, in which participants learn to predict effects on the basis of potential causes, constitute the classic paradigm in the study of causal induction. Learning that a specific cold medicine may cause an allergic rash is a typical example of predictive learning. Predictive learning is central to our gaining knowledge about the causal consequences of observed events and allows us to predict or control future events. Diagnostic learning, in which participants learn to diagnose causes on the basis of information about effects, is one of several other types of causal induction that are often not treated in extant theories. When, for example, we wake up with high fever and find out later that the symptom was probably caused by eating tainted food, we engage in diagnostic learning. Despite the fact that predictive and diagnostic learning clearly differ, few theories of causal induction model them differently.

Competing Theories of Causal Learning

The Associative View

A number of researchers have recently claimed that causal learning is a special case of associative learning (e.g., Gluck & Bower, 1988; Shanks & Dickinson, 1987). Associative learning theories can be characterized by two basic assumptions, which can be separated although they tend to come in tandem. First, associationistic theories assume that the learning experience can be represented solely in terms of two types of event representations, cues and outcomes. Cues are events that occur temporally prior to outcomes and play the role of eliciting responses; outcomes are the events to which the responses refer. Cue and outcome representations are linked by associative weights. In learning tasks, feedback is typically given about the outcome that should have been predicted. Due to this reduction of learning to acquiring associations between cues and outcomes predictive and diagnostic learning are conceived of as identical learning tasks, provided that cues and outcomes are kept constant (see also Waldmann, 1996; Waldmann & Holyoak, 1990, 1992; Waldmann, Holyoak, & Fratianne, 1995). In predictive tasks, cues correspond to causes and outcomes to effects that are predicted; in diagnostic learning, cues correspond to effects and outcomes to causes that are diagnosed. According to associative accounts that endorse the cue-outcome framework, if participants in a learning experiment are presented with information about the presence or absence of substances in people's blood as cues for the classification of a disease, it does not matter whether these substances represent causes or effects of the disease. As long as cues and outcomes are identical, learning and the ensuing mental representation should be identical.

The second assumption of most associationistic theories is that learning is based on a learning rule that modifies the weights of the associative links between cues and outcomes on a trial-by-trial basis (e.g., Rescorla & Wagner, 1972). The exact nature of the postulated learning rule differs from theory to theory. This article will focus mainly on the Rescorla–Wagner rule, which is the most widely postulated in tasks involving causal induction. It should be noted, however, that the foregoing results showing differences between predictive and diagnostic framings of otherwise identical tasks are inconsistent with any theory that reduces

Most of the research reported in this article was conducted while Michael R. Waldmann was affiliated with the Department of Psychology of the University of Tübingen and the Max Planck Institute for Psychological Research, Munich, Germany. The research was in part funded by a grant from the Deutsche Forschungsgemeinschaft (Wa 621/5-1,2).

I would like to thank V. Chase, K. Holyoak, B. Malt, L. Martignon, U. Reips, and B. Whitlow for helpful comments. Special thanks to M. Seemann who helped with the mathematical appendix.

Correspondence concerning this article should be addressed to Michael R. Waldmann, Department of Psychology, University of Göttingen, Goßlerstraße 14, 37073 Göttingen, Germany. Electronic mail may be sent to michael.waldmann@bio.uni-goettingen.de.

learning to the acquisition of associations between cues and outcomes, regardless of the learning rule it postulates.

The Causal-Model View

This view holds that participants constrain the induction process by imposing causal models on their observations when learning about new causal relations (see Waldmann, 1996; Waldmann & Martignon, 1998). Whereas associationist theories assign events to internal representations solely on the basis of temporal order, initial events being cast as "cues" and later events as "outcomes," causal-model theory maps events to mental models on the basis of their causal role in the physical world. Cues that play the role of potential causes yield structurally different expectations from cues that play the role of potential effects (see also Connolly, 1977; Connolly & Srivastava, 1995; Tversky & Kahneman, 1980). Thus, causal models contain information about causal directionality: Causes influence effects and not the other way around.

Knowledge about the asymmetry of causes and effects is central for our ability to act in an appropriate way. For example, if a substance is the cause of a disease, the disease occurs when the substance is given to an organism; however, producing the effect, the disease, by different means does not cause the presence of this substance. Moreover multiple causes of a common effect converge, whereas multiple effects of a common cause diverge (Reichenbach, 1956). For example, two switches that are connected to the same light in some way collaborate in their causal influence on the light. In contrast, a light switch that is connected to two lights turns on the two lights independently. Thus, the deeper issue behind the debate between associative theories and causal-model theory is whether humans are sensitive to one of the most fundamental features of causes and effects in the physical world, namely, their asymmetry, or whether they reduce learning events to cues and outcomes, which can give rise to mental representations that contradict physical reality.

The causal-model view also rejects the assumption of associative theories that causal strength is represented by associative weights. It holds instead that humans, as well as some animals, are able to use more sophisticated learning rules that make use of representations of frequencies, conditional probabilities, and contingencies. This claim is based on substantial evidence that humans and animals represent frequency and contingency information (Cosmides & Tooby, 1996; Gallistel, 1990; Hasher & Zacks, 1979). In the case of binary discrete causes and effects, causal-model theory claims that people assess whether the presence of the causes increases or decreases the probability of the effects. In situations in which it is not necessary to control for co-factors (e.g., a single cause that has a single effect or common-cause models in which a single cause has multiple, independent effects), causal strength may be estimated by computing the unconditional contingency between the cause and its effect, $p(E|C) - p(E|\sim C)$, with the two components of the formula representing the probability of an effect conditional upon the presence and the absence of the cause, respectively.[1] In situations in which multiple causes converge on the same effect (i.e., common-effect models), it is necessary to control for the cofactors when assessing the causal strength of a specific cause. If, for example, we plan to assess the hypothesis that smoking (C) causes heart disease (E), we must control for alternative causes, such as eating junk food (K). This can be accomplished by computing conditional contingencies separately for the subset of people who eat junk food (K) and for the subset of people who do not eat junk food ($\sim K$). If it turns out that the cause alters the probability of the effect in both subgroups to the same extent, it can be concluded that smoking is an independent causal factor (see also Melz, Cheng, Holyoak, & Waldmann, 1993).

Blocking in Predictive Versus Diagnostic Learning

Since Kamin (1969) discovered the phenomenon of blocking in animal learning, cue competition has been a basic phenomenon addressed by virtually all associative learning theories. In Phase 1 of the classic blocking paradigm, animals are trained to associate an initial conditioned stimulus (CS_1) with an unconditioned stimulus (US). For example, they may learn to predict a shock outcome (US) on the basis of a tone cue (CS_1). In Phase 2 of the learning procedure, a second cue (CS_2), for example, a light, is redundantly paired with the initial tone cue. Kamin's crucial finding was that, in spite of being perfectly correlated with the outcome, the redundant light cue did not seem to acquire any associative strength for these animals as compared to a control group that did not receive Phase 1 training.

According to Rescorla and Wagner's theory (1972), blocking results from a failure of the CS_2 to acquire associative strength. Because the CS_1 learned in Phase 1 allows for perfect predictions in Phase 2, no further learning occurs. Alternative associationistic accounts attribute blocking to shifts in selective attention or to processes in the response generation phase (e.g., Mackintosh, 1975; Miller & Matzel, 1988; Pearce & Hall, 1980). However, these theories likewise categorize learning events as cues and outcomes and therefore predict blocking of the redundant CS_2, regardless of whether the cues represent causes or effects.

Competition in Causal Models

According to associative theories, blocking is a consequence of the fact that cues may represent information redundant with cues that were presented earlier. By contrast, according to causal-model theory, the primary basis of blocking and other types of competition are the causal relations represented in causal models, not temporal order of the learning input. Causal-model theory does not predict

[1]Cheng (1997) has recently proposed a different measure of causal strength. The present article does not aim at distinguishing between contingency theories and Cheng's causal-power theory. In all experiments deterministic causal relations were used for which Cheng's measure and contingency measures provide identical estimates.

competition among cues, it predicts potential competition among causes in common-effect structures.

For common-effect structures, it is appropriate to hold co-factors constant when assessing the causal strength of the individual causes. In predictive blocking designs with the cues representing causes of a common effect, it is typically impossible for learners to hold the predictive cue constant when the relation between the redundant cue and the outcome is to be assessed. That is, the redundant cue is never presented in the absence of the predictive cue, and the conditional contingency between the redundant cue and the outcome in the absence of the predictive cue is therefore undefined. Furthermore, the predictive cue already causes the effect with the maximal probability of 1, which creates a ceiling situation that makes it impossible for learners to observe whether the redundant cue has any impact above the influence of the predictive cue (see Cheng, 1997). Thus, causal-model theory predicts that in such a deterministic common-effect situation, participants should be uncertain as to whether the redundant cue is an independent cause of the outcome or not. In contrast to the Rescorla–Wagner theory, causal-model theory predicts that blocking in this case will be partial. Participants will be uncertain about the status of the redundant cue rather than certain that it is not a cause. By contrast, in a common-cause situation, causal-model theory predicts no competition among independent effects of a common cause. Because there is only one cause in common-cause situations, no co-factors need to be held constant. In this case, it is appropriate for participants to use the unconditional contingencies between this cause and its effects as a measure of causal strength.[2]

Predictive Versus Diagnostic Inferences

Knowledge about causal structures, such as common-cause or common-effect models, can be acquired either in the predictive (cause–effect) or diagnostic (effect–cause) input order. Independent of the input order the test questions can also be framed in either the predictive or the diagnostic direction. For example, a predictive test question might ask about the probability of a specific symptom (effect) when a disease (cause) is present, whereas the corresponding diagnostic test question would ask about the probability of the disease (cause) when the symptom (effect) is present.

Causal-model theory predicts that, at least with relatively simple causal structures, people should be aware of the difference between predictive and diagnostic test questions (Waldmann & Holyoak, 1992). An important feature of diagnostic inferences is the necessity of taking into account alternative causes of the observed effect. Even though fever may be a deterministic effect of a flu, it is nevertheless a bad diagnostic sign of flu because fever has many alternative causes. Thus, a symptom such as fever should yield high predictiveness ratings if the task requires a cause–effect rating (from flu to fever) but low ratings if the task requires the effect–cause inference (from fever to flu). Whereas the causal strength of a cause–effect relation is not affected by other collateral effects, the strength of the corresponding effect–cause relation is dependent on whether other causes also might produce the effect. In summary, the predictions of causal-model theory are based jointly on assumptions about the causal models underlying the learning events and the type of test questions that access these causal models.

Waldmann and Holyoak (1992) used a variant of the two-phase blocking design to test associative theories against causal-model theory. In their Experiment 3, for example, participants received trials with information about the state of buttons on a computer screen (*on* or *off*) and were requested to express whether they believed that the alarm connected to the buttons was on or off (*yes* or *no*) in a given trial. Feedback was given after each decision. In Phase 1, they learned that one button, the predictive cue, was a deterministic predictor of the state of the alarm. In Phase 2, this predictive cue was constantly paired with a redundant cue. Both buttons were either on followed by the alarm being on, or both were off and the alarm also was off. All participants saw identical stimuli.

The key manipulation involved the causal interpretation of the cues and the outcome. The initial instructions characterized the cues either as potential causes of a common effect (common-effect model) represented by the outcome, or the very same cues were described as potential effects of the outcome, which in this condition represented a common cause (common-cause model). Thus, in one condition the buttons were described as potential causes of the state of the alarm, and in the other condition the buttons were characterized as effects whose state was determined by the state of the alarm.

Participants in both conditions were asked, in identically phrased test questions, to assess how predictive each button was for the state of the alarm. To permit the use of identical test questions in the two learning conditions, the questions did not mention the causal status of the cues and outcomes. Thus, whether learning and the test questions referred to a predictive or a diagnostic situation was solely manipulated by means of the initial cover stories. It was predicted that the two cover stories would lead participants to form causal models in which cues and outcome were assigned different causal roles (i.e., of causes or effects).

Because participants in both conditions received identical cues, identical outcomes, and identical test questions, associative theories predict identical learning in this experiment. In particular, most theories (e.g., the Rescorla–Wagner theory) would predict blocking of the redundant cue. By contrast, causal-model theory predicts competition in the common-effect but not in the common-cause situation. According to causal-model theory, the cover stories should lead participants in the common-effect condition to form a causal model in which the cues (buttons) are assigned the status of potential causes and the outcome (alarm) the status

[2]It is important to note that common-cause structures are not the only way multiple effects can be linked. There may be complex causal networks underlying the observed events (see Pearl, 1988; Waldmann, 1996). Causal-model theory does not generally predict that effects do not compete, it only makes this prediction for situations in which these effects are conditionally independent given a common cause (and for structurally similar situations).

of an effect. Since both learning and the test questions are directed from cues to outcomes both processes are directed in the predictive cause-effect direction within the causal model. In this situation, causal-model theory predicts competition among the cues representing causes. Participants should be uncertain about the causal status of the redundant cue. In the contrasting condition, the initial common-cause instructions should lead to a causal model in which the cues (buttons) represent effects and the outcome (alarm) the common cause. As both learning and the test questions again are directed from cues to outcomes both processes are directed from effects to causes within the causal model participants presumably impose on the learning input. For this situation causal-model theory predicts equal ratings of the predictiveness of the predictive and the redundant cue. Participants should have learned that the alarm is a deterministic cause of both effects. Furthermore, since no alternative causes of the states of the buttons are mentioned that could lower the diagnostic validity of the cues, no differential lowering of the diagnostic ratings is expected.

The results clearly supported causal-model theory. Whereas the redundant cue was rated significantly lower than the predictive cue in the predictive common-effect condition, no reliable difference was observed in the diagnostic common-cause condition. In Experiment 2 Waldmann and Holyoak (1992) also provided evidence for the predictions concerning diagnostic inferences. Even though participants again learned about a situation with a deterministic cause of two effects, the diagnostic ratings of the redundant effect (being underweight) were lowered. Apparently, participants were sensitive to the fact that there are many alternative causes of this symptom and lowered their diagnostic ratings accordingly.

Criticisms of the Causal-Model Account of Cue Competition

Waldmann and Holyoak's (1992) causal-model theory and the demonstrations of asymmetries of cue competition provoked a number of responses from proponents of the associative view (also Waldmann, 1996; Waldmann & Holyoak, 1997, for responses to some of the criticisms). Some theorists have argued that associationist theories are perfectly able to handle these results. Others acknowledge that the results are potentially problematic for associative learning theories, but argue that they are not convincing enough to warrant giving up associative learning theories. These two responses will be discussed in turn.

Response 1: Associative Learning Theories Predict Asymmetry of Cue Competition

Van Hamme, Kao, and Wasserman (1993) pointed out that the Rescorla–Wagner rule has a built-in asymmetry between cues and outcome that may underlie the observed asymmetry of causes and effects. According to this learning rule, cues compete to predict the common outcome, but multiple outcomes of individual cues do not compete with each other. Thus, in learning situations in which cues represent causes and outcomes effects, the Rescorla–Wagner rule predicts competition among causes but not among effects. This asymmetry between causes and effects has been firmly established in a number of experiments with animals and humans (e.g., Baker & Mazmanian, 1989; Baker, Murphy, & Vallée-Tourangeau, 1996; Matute, Arcediano, & Miller, 1996, Experiments 1, 2; Rescorla, 1991, 1995; Van Hamme et al., 1993). All these studies have in common that the causes were either presented prior to their effects as cues, or causes and effects were presented simultaneously so that causes could be assigned the role of cues within an associative network.

The critical test case for distinguishing between associative and causal-model theory, however, presents diagnostic learning situations in which the effects are presented as cues prior to their causes (as in Waldmann & Holyoak, 1992). In such situations the Rescorla–Wagner rule predicts competition among effects but not among the causes, a pattern contrary to physical relations in the world and contrary to the results of Waldmann and Holyoak's (1992) experiments. Thus, the fact that the Rescorla–Wagner rule sometimes makes the right predictions is not due to the fact that it conceptually distinguishes between causes and effects as causal-model theory does; rather it is a consequence of a fortuitously valid mapping between the learning rule and a specific set of learning situations.

Van Hamme et al.'s (1993) observation of the asymmetry of the Rescorla–Wagner rule may also be read as a general suggestion to map causes to the input level and effects to the output level regardless of input order. However, this proposal faces the problem that it is unclear how diagnostic-learning tasks are mastered when effects are presented as cues prior to feedback about the outcome (e.g., when discovery of one's fever precedes discovery that one has eaten tainted food). It is not clear how a network that maps effects to the output level would generate a response when presented with effect information as the input to a diagnostic decision. It is also unclear how a network that acquires single associative weights between each cue and the outcomes could explain people's ability to differentiate between predictive and diagnostic inferences (e.g., Waldmann & Holyoak, 1992).

To overcome this problem, Shanks and Lopez (1996) proposed a more complex theory of associative learning. According to this theory two associative networks may be run in parallel, one directed from causes to effects and the other directed from effects to causes. The latter network is supposed to handle diagnostic learning and diagnostic inferences. This modified theory explains the absence of blocking in Waldmann and Holyoak's (1992) Experiment 1, in which participants from the common-cause condition gave predictive cause-effect ratings after diagnostic learning. It is, however, refuted by the absence of a significant blocking effect in the diagnostic condition of Experiment 3, in which both the learning and the inferences requested in the test phase were diagnostic (i.e., from effects to causes).

Response 2: There Is No Asymmetry of Cue Competition

Some critics of causal-model theory questioned the data reported in Waldmann and Holyoak (1992) rather than trying to modify the standard associationist framework to accommodate it. In particular, they dismissed the most problematic result for associative learning theories, namely, the absence of blocking in the diagnostic condition of Experiment 3 (Waldmann & Holyoak, 1992) after diagnostic learning and diagnostic test questions. Matute et al. (1996) suspected that this finding might not be replicable. Some proponents of the associative view argued that lack of statistical power may have prevented the small observed descriptive difference between the ratings of the predictive and the redundant cue (i.e., the blocking effect) from becoming statistically significant (Matute et al., 1996; Shanks & Lopez, 1996). These critiques downplay what, according to Waldmann and Holyoak (1992), was the most important result: a highly reliable interaction between the causal status of the cues and blocking. In addition, the generality and validity of the findings were called into question (Matute et al., 1996; Shanks & Lopez, 1996), and it was suggested that the two-phase blocking paradigm used might favor nonassociative types of learning (Price & Yates, 1995). Finally, Waldmann and Holyoak's assumption that prior knowledge caused participants in their Experiment 2 to give low diagnostic ratings to an effect that suggests alternative causes (being underweight) was sometimes misinterpreted or ignored, which opened up the possibility of viewing the obtained low ratings of the redundant cue as evidence for blocking and the results of Experiment 3 (in which prior knowledge was excluded) as a mere failure to replicate (Matute et al., 1996; Miller & Matute, 1998).

Finally, a number of studies have been reported that exhibit competition among effects, a finding that contradicts the claim of causal-model theory that competition should be observed among alternative causes of a common effect but not among effects of a common cause (Matute et al., 1996; Price & Yates, 1995; Shanks & Lopez, 1996). For example, Shanks and Lopez (1996) reported results that seemed to demonstrate cue competition regardless of whether the cues represented causes or effects. However, as pointed out by Waldmann and Holyoak (1997), these experiments were not entirely convincing: It is questionable whether the cover stories conveyed clear causal interpretations and whether the test questions properly assessed causal knowledge. Most importantly, Shanks and Lopez's hypothesis that causal order did not affect learning was not directly tested because they did not experimentally manipulate the causal role of the cues. Finally, a statistical reanalysis of some of the results produced results that favor causal-model theory. Other problems with apparent refutations of causal-model theory will be discussed in the General Discussion (see also Waldmann, 1996; Waldmann & Holyoak, 1997).

Overview of Experiments

The main goal of the present experiments is to provide further evidence bearing on the debate between the associationist and the causal-model account of blocking. The most disputed prediction of causal-model theory is that there will be no competition among effects in diagnostically acquired common-cause structures with diagnostic test questions. Experiment 1 replicates this finding and reveals a difference between predictive and diagnostic learning using a novel design that takes care of some of the criticisms raised against the experiments in Waldmann and Holyoak's (1992) article (Matute et al., 1996; Shanks & Lopez, 1996). In particular, some critics argued that the comparison between the predictive and the redundant cue is confounded with different numbers of trials. Experiment 1 uses a design in which the test of the blocking effect was based on cues that were presented an equal number of times.

Experiment 2, which focuses on the diagnostic condition, tests assumptions of causal-model theory that previously have not been made explicit. It will be shown that the predicted absence of the blocking effect hinges on the learners' retrospective assumptions about the presence of the new redundant effect in the previous learning phase, Phase 1, in which they had not yet learned about this effect. These retrospective assumptions are invited by the structure of common-cause models, and are therefore a side effect of sensitivity to causal directionality. This analysis may also explain why a small difference between the predictive and the redundant cue has sometimes been observed (e.g., Waldmann & Holyoak, 1992, Experiment 3).

Experiment 3a directly tests the assumption of causal-model theory that diagnostic inferences are sensitive to the presence of potential alternative causes. Critics have correctly pointed out that Waldmann and Holyoak (1992) tested this assumption only indirectly on the basis of a cross-experiment comparison in which different materials were used. One of the diagnostic conditions of Experiment 3a also serves as an additional test for causal-model theory's most disputed prediction: the absence of blocking after diagnostic learning and diagnostic test questions when no alternative causes are available. Finally, this experiment also includes predictive-learning conditions that test additional predictions of causal-model theory against associative accounts.

Experiment 3b focuses on the predictive-learning condition of Experiment 3a. Experiment 3b varies the length of training in Phase 1 to test causal-model theory's account of the predictive condition of Experiment 3a against an associationist alternative explanation of the results, which invokes preasymptotic learning as a possible reason for the obtained patterns.

Experiment 1

The goal of this experiment is to demonstrate the asymmetry of predictive and diagnostic learning using a task context and a blocking paradigm different from the one used by

Waldmann and Holyoak (1992). Some critics have argued that a statistical comparison between the predictive cue from Phases 1 and 2 and the redundant cue from Phase 2 is not sufficient to establish a blocking effect as the number of presentations of these cues is not kept constant (Matute et al., 1996; Shanks & Lopez, 1996). Thus, different ratings of these two cues may be due to the differences of trial numbers rather than cue competition. This argument does not account for the reliable interaction between the predictive and the diagnostic learning condition with identical learning trials; a significant difference of the ratings was only observed in the predictive but not in the diagnostic learning condition. However, it still seems desirable to test the asymmetry hypothesis with a design that keeps the number of trials constant.

The present experiment uses a design that is adapted from recent studies by Chapman and Robbins (1990) and Williams, Sagness, and McPhee (1994). In these experiments a predictive learning task was used in which the causal effect of stocks on the market had to be predicted. In Phase 1, participants learned, for example, that Stock P predicted a market change (i.e., P+), whereas Stock N, the nonpredictive cue, did not predict a change (i.e., N−). In Phase 2, Stock R, the redundant cue, constantly accompanied Stock P followed by a market change (PR+). The market also changed when Stock N was paired with Stock I, the informative cue (NI+). The Rescorla–Wagner rule and other associative theories predict blocking of Stock R by the predictive cue, Stock P. The advantage of this paradigm is that the blocking effect can be tested by comparing the two cues Stock R and Stock I. Both cues have been presented the same number of times followed by the same outcome. Furthermore, they both were only presented as components of compound cues (PR or NI). Thus, a lowering of the rating of Stock R relative to Stock I indicates the impact of the predictive Stock P versus the nonpredictive Stock N on the ratings.

Williams et al. (1994) found that in many learning situations this paradigm did not yield a significant blocking effect (i.e., Stocks R and I were rated equivalently), which, according to the authors, indicates that learners tended to treat the compounds as novel configural cues and not as conjunctions of elemental cues. However, Williams et al. also showed that prior experience that encourages elemental processing (Experiment 4) or instructions that encourage viewing the stocks as separate entities (Experiment 5) yielded a reliable blocking effect consistent with the previous findings of Chapman and Robbins (1990).

The present experiment modifies this paradigm to permit the use of identical trials in both predictive and diagnostic learning contexts. The problem with the original paradigm mainly arises in the diagnostic condition, which can be best seen with a concrete example. Consider a diagnostic cover story in which the cues represent effects of a common cause. For example, the presence of different symptoms (cues) may be caused by a specific novel disease (outcome). Adopting the outlined paradigm (Chapman & Robbins, 1990; Williams et al., 1994) would then entail a causal structure in which participants learn within Phase 1 that the disease causes Symptom P (P+) but does not produce Symptom N (N−). In Phase 2, learners would then observe that the disease causes the compound of Symptoms P and R (PR+) as well as the compound of Symptoms N and I (NI+). These observations are incompatible with common-cause models with stable causal characteristics because they signify a change of the causal power of the cause between the two learning phases: In Phase 1 the disease never causes Symptom N, whereas in Phase 2 Symptom N is always seen when the disease is present. The goal of Experiment 1 was to present identical learning input that is equally compatible with the causal models suggested in the predictive condition (common-effect model) and the diagnostic condition (common-cause model). Therefore, the paradigm had to be slightly modified.

In the experiment participants were instructed that they were going to learn about a fictitious box. They were going to receive information about the state of lights on the front side and had to predict the state of one light on the invisible back side. In the predictive condition (common-effect condition) the visible lights were characterized as indicators of potential causes and the light on the back side was the effect, whereas in the diagnostic condition (common-cause condition) the lights on the front side represented effects, and the light on the invisible back side the potential cause of these effects. Apart from these differential initial cover stories the learning trials were identical across both conditions. In Phase 1 participants learned that one of the four lights on the front side was predictive of the outcome, the light on the back side (i.e., P+). Information about the state of the other cues was withheld within this Phase. In Phase 2 this predictive light was constantly paired with a second, redundant light (i.e., PR+) followed by the outcome. Alternately, a second compound of two lights, two informative cues, was shown that also was predictive of the outcome (i.e., II+). Thus, one informative cue was presented along with a second, redundant informative cue (as opposed to being paired with a non-predictive cue). A blocking effect is indicated by a lowering of the rating of the redundant cue (R) relative to the two other new cues of Phase 2 (I) which have been presented the same number of times. Thus, like in the studies of Chapman and Robbins (1990) and Williams et al. (1994), blocking could be measured by comparing the ratings of cues that were presented an equal number of times within compounds.

Because all participants received identical learning inputs, associative theories that model learning as the acquisition of associations between cues and outcomes predict identical learning. In particular, the Rescorla–Wagner theory and many other associationist theories predict a blocking effect in both conditions signified by a reliable difference between the redundant cue and the informative cues. By contrast, causal-model theory predicts an interaction. A blocking effect should be observed in the predictive condition (i.e., lowering of the redundant cue relative to the informative cues) but not in the diagnostic condition.

In the predictive condition, participants learn that the predictive cue on the front side is a deterministic cause of the light on the back side. Thus, in Phase 2 they should be

uncertain about the status of the redundant cue as it is never shown in the absence of the predictive cue. They should become aware of the fact that whatever the causal status of the redundant cue is, the observed outcome is definitely simultaneously caused by the predictive cue. Thus, the possibilities for the redundant cue range from overdetermining the effect to not having a causal impact at all. The whole range is compatible with the observed learning events. By contrast, the two other new cues (I) are presented as an alternative compound cause of the outcome. Thus, it also is not possible to assess the individual causal impact of each individual cue in the absence of the collateral cue. Overdetermination also is possible but attributing low causal power to one cue would imply the attribution of the complementary high causal power to the other cue. Lacking a way to discriminate between the cues, a parsimonious assumption is to equally divide the causal power between the two cues. Thus, it is reasonable to expect higher ratings of the two new cues (I) than for the redundant cue (R) in the predictive context. This finding is also predicted by the Rescorla–Wagner rule. No difference is predicted by theories that postulate configural processing of the compounds (Williams et al., 1994).

The predictions diverge in the diagnostic context. Again the Rescorla–Wagner theory predicts blocking of the redundant cue by the predictive cue. By contrast, causal-model theory predicts high and equivalent ratings for all cues. In Phase 1, participants learn that the effect light on the front side (predictive cue) is causally influenced by the light on the invisible back side. Again information about the state of the other lights is not given. In Phase 2, participants alternately are informed about the two compounds (PR+ or II+), while information about the nonpresented compound is withheld (as in the predictive condition). Thus, participants should eventually infer that apparently the cause light deterministically influences all four lights. Therefore, unlike most associative theories, causal-model theory does not predict a blocking effect in this condition.

Method

Participants and design. Participants were 24 students from the University of Munich, Germany, who were randomly assigned to the diagnostic or predictive learning conditions (12 participants in each condition).

Procedure and materials. Participants were run individually on a microcomputer. Prior to the learning task they received written instructions (in German). In the predictive-learning condition (common-effect model) the instruction stated that the task would be to learn about causal relations. Participants were supposed to imagine a box with four lamps on the front side and one lamp on the back side. Only the front side but not the back side could be seen from the perspective of the learner. Switches were attached to the lamps of the front side that switched on the respective visible light. The task was to learn to predict whether the switch also turned on the light on the back side of the box. Furthermore the instructions stated that during the learning task information about the current state of the lights on the front side would be given (*on* or *off*). Whenever a light was on, participants should imagine that the experimenter had switched on the light.

In the diagnostic-learning condition (common-cause model) similar instructions were given. The only difference was that n-switches for the four visible lights on the front side were mentioned, and that the lights were characterized as potential effects of the light on the invisible back side. Now the light on the back side was described to contain a switch that the imaginary experimenter occasionally, invisible to the learner, turned on or off. Again it was stated that participants were going to receive information about the states of the visible lights on the front side, and that they were expected to judge whether, as a consequence of the presumed actions of the experimenter, the light on the back side was also on or off. Thus, in the diagnostic condition the four lights on the front side represented potential effects of the common-cause light on the invisible back side, whereas in the predictive condition the four lights on the front side represented potential causes of a common-effect light on the back side.

After these initial instructions participants were requested to summarize the instructions to make sure that they understood the causal structures conveyed in the different cover stories. Then the learning trials on the computer commenced. All participants received identical learning trials with identical learning instructions. These instructions stated that participants were going to receive information about the state of four colored lamps on the visible front side. These lamps were either on or off. Furthermore it was mentioned that instead of this information four question marks may also be shown which indicate that the current state of the lamp cannot be seen during the respective trial. The task was to say "yes" when the participants believed that the light on the back side also was on, and say "no" when it was presumably off. After their decision they would be given immediate feedback. Then participants were alerted that later they would be asked about the different lamps so that it was useful to memorize the positions of the colors.

The learning trials showed a screen with the header "front side" above the four capitalized color names red, green, white, and blue next to each other on one line. In Phase 1, only information about one light, the predictive cue, was given; the other three lights were marked with question marks (e.g., "RED: ON GREEN: ???? WHITE: ???? BLUE: ????"). Thus, this example indicates that the red light on the left side is on, whereas the current state of the other three lights is unknown. The correct answer was to say "yes" when the predictive light was on and "no" when it was off (i.e., P+). After each decision the experimenter hit one key, which displayed a screen with the feedback. The feedback screen showed the header "back side" on top and below information about the state of the lamp on this side (e.g., "Lamp: ON"). The learning task was extremely simple so that all participants learned it within six trials (three "yes" trials, three "no" trials). For half of the participants, the red light on the left side was the predictive cue; for the other half, the blue light on the right side was.

After this learning phase participants were requested to rate the predictiveness of the light they had seen using a number between 0 (*you are certain that the light on the back side is off*) and 100 (*you are certain that the light on the back side is on*). The rating instructions stated that the participants should imagine, for example, that the blue light on the front side was on. Then they should judge how well this light by itself predicts the state of the light on the back side of the box.

In Phase 2 of the learning procedure four different trial types were presented in which information about two lights was given while the other two lights were marked with question marks. Two trial types consisted of pairing the predictive light from Phase 1 with a new redundant light (PR+). Either both lights were off or both lights were on (e.g., "RED: ON GREEN: ???? WHITE: ON BLUE: ????"). The other two trial types (II+) presented the two other lights either both being on or both being off (e.g., "RED: ???? GREEN: OFF WHITE: ???? BLUE: OFF"). Whenever the lights

were on, the light on the back side also was on ("yes"); otherwise it was off ("no"). These patterns were presented three times each in a random order. The assignment of the redundant cue to one of the three remaining lights was counterbalanced across participants. After the learning phase participants rated the predictiveness of the four lights.

Results and Discussion

The results clearly support causal-model theory. After the six learning trials of Phase 1 all but two participants gave maximal ratings (i.e., 100) to the predictive light. The two participants chose the ratings 80 (predictive-learning condition) and 50 (diagnostic-learning condition). Figure 1 shows the mean ratings obtained after Phase 2. The rating patterns after Phase 2 differed between the two conditions. In the diagnostic condition every single participant gave identical ratings to the four lights, thus preempting any need for statistical analysis of the blocking effect in the diagnostic condition. In fact all participants except for one (who gave all lights the rating of 50, see above) rated the lights with the maximal number 100. Thus, the results show no indication whatsoever of a blocking effect, which clearly speaks against the predictions of the Rescorla–Wagner rule and many other associative theories.

By contrast, a clear blocking effect was observed in the predictive condition, which shows that configural-cue theories also do not account for the results of this experiment. The most crucial empirical indicator of this effect involved a comparison between the redundant cue and the two other (informative) cues. (As all but one participant rated these two cues identically the statistical analyses are based on the average of these cues.) As can be seen in Figure 1, the redundant cue was rated clearly lower than the informative cues, $F(1, 11) = 10.6, p < .01, MSE = 726.9$. Nine out of 12 participants in the predictive-learning condition decided to give lower ratings to the redundant cue than to the informative cues. Both cue types had been presented an equal number of times as components of compound cues. The asymmetry of cue competition is also indicated by the difference between the predictive cue and the redundant cue in Phase 2, $F(1, 11) = 37.3, p < .001, MSE = 835.1$, as the lack of a difference of the ratings within the diagnostic condition shows that this difference cannot be attributed to different numbers of learning trials.

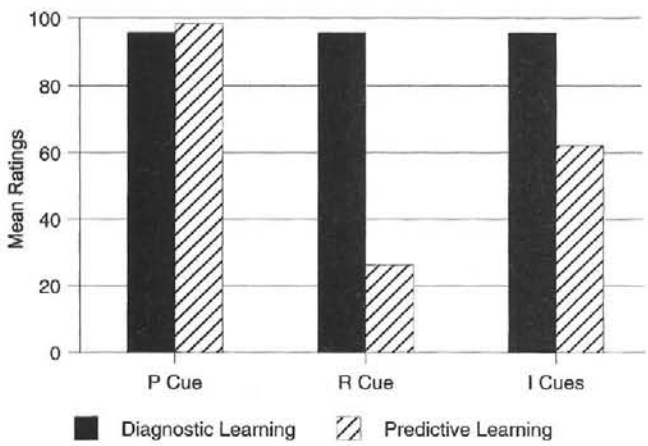

Figure 1. Mean predictiveness ratings from the diagnostic and predictive conditions in Phase 2 of Experiment 1, for the predictive cue (P cue), the redundant cue (R cue), and the average of the informative cues (I cues).

Experiment 2

Perhaps the most disputed finding of Waldmann and Holyoak (1992) is the absence of a blocking effect in the diagnostic condition of their Experiment 3. Whereas the present Experiment 1 replicated this effect using a different paradigm, Experiment 2 explores some boundary conditions of this effect that have been left implicit in the exposition of causal-model theory in Waldmann and Holyoak (1992).

A basic variant of the diagnostic blocking paradigm was used in this experiment. In Phase 1 of learning, a single effect cue (predictive cue) is established as a perfect diagnostic cue for its cause (i.e., P+). In Phase 2, a second, redundant effect cue (redundant cue) is constantly paired with the predictive cue (i.e., PR+). Within this phase both cues are perfectly correlated with the cause.

Causal-model theory predicts the absence of cue competition when the participants view the two effect cues as independent effects of a common cause. In this situation the effects are independent of each other conditional upon their common cause so that knowledge about a causal relation between the cause and the first effect should not influence learning about the causal relation between the cause and the second effect. In common-cause models the causal relation between the cause and each effect should be assessed using unconditional contingencies between the cause and each effect. Thus, provided that the potential influence of alternative causes is kept constant for both effects, they should be identically rated when the unconditional contingencies are identical.

A closer look at the two-phase blocking paradigm reveals that the unconditional contingencies for the predictive and the redundant cue only are identical when specific boundary conditions hold. Participants see a maximal unconditional contingency (of 1) between the outcome (cause) and the predictive cue (Effect 1) throughout both learning phases. However, participants receive information about the redundant cue only during Phase 2. Within this learning phase the contingency is also maximal (=1). However, whether Phase 2 is viewed as representative for the unconditional contingency between the outcome and the redundant cue (Effect 2) across both learning phases is dependent on retrospective assumptions about the presence or absence of the redundant cue during Phase 1 in which information about this cue had been withheld from participants. If the redundant cue is retrospectively viewed as partly absent in Phase 1, the unconditional contingency between the cause and this cue, that is Effect 2, would be lowered relative to the unconditional contingency of the cause and Effect 1. Retrospectively, cases would be integrated in the contingency assessment in which the cause was present but Effect 2 absent. In

contrast, if learners assume that they were not informed about the status of the redundant cue they may be led by the common-cause model to infer that Effect 2 was present all along when the cause was present. This retrospective inference is invited by the assumption that the causal power of the common cause probably is stable across learning phases.

Waldmann and Holyoak's (1992) experiments clearly differentiated between explicitly absent cues and cues that were not mentioned (as in the present Experiment 1 in which "????" represented lack of information). For example, in their Experiment 3 one of the three lights that played the role of potential effects of a common cause was explicitly described as being off in all trials of both learning phases. In contrast, the light that played the role of the redundant effect cue was not mentioned before Phase 2 started. Thus, it was consistent with the presentation of the trials that this second effect light (redundant cue) also was on whenever the predictive light (predictive cue) was on in Phase 1; it was simply not mentioned.

Common-cause models invite this backward inference that the redundant cue has already been an effect of the cause during Phase 1. In Phase 2, participants learned that the cause has a second deterministic effect. Because in Phase 1 this second effect had not been mentioned, it was plausible to assume that it had been an effect of the common cause all along. Alternatively, the participants would have had to make the rather implausible assumption that the cause has changed its capacity from causing one effect to causing two effects in the course of the two learning phases.

Because common-cause models with stable causal characteristics imply this backward inference, no difference in the assessments of the predictive and the redundant cue was expected. However, some participants may have interpreted the absence of information about the redundant cue in Phase 1 as implying the absence of this cue during this phase. For these participants, causal-model theory would predict a difference between the ratings for the predictive and the redundant cue as they should estimate the unconditional contingency of the redundant cue to be lower than that of the predictive cue across both phases. Furthermore, they should be hesitant about the relation between the cause and the second effect, as the cause has apparently changed its causal power with respect to this effect. This may explain the small (nonsignificant) differences between the predictive and the redundant cue in the diagnostic conditions of the experiments of Waldmann and Holyoak (1992). It should be noted that this difference is not hypothesized to be a result of cue competition; rather, it is a natural consequence of differences of the causal power of the common cause with respect to its two effects.

To test the assumption that the observation of a difference between the predictive and the redundant cue in the diagnostic learning condition is dependent on whether the redundant cue is viewed as being absent throughout Phase 1, Experiment 2 compares two conditions. In the no-information condition, a condition that corresponds to the previously used diagnostic learning tasks, participants were not informed about the status of the redundant cue in Phase 1, whereas in the explicit-absence condition participants received explicit information about the absence of the redundant cue during this phase. (Experiment 2 also includes an additional uncorrelated cue that is constantly absent throughout all trials.)

Causal-model theory predicts similar ratings of the predictive and the redundant cue in the no-information condition, which is consistent with the assumption that the causal power of the common cause remains stable across the two learning phases. In the explicit-absence condition, the ratings for the redundant cue should be considerably lower than for the predictive cue. The information about the absence of the redundant cue during Phase 1 lowers the unconditional contingency between the cause and this cue. This apparent change of causal power should be reflected in the ratings. With additional training in Phase 2 these differences should, however, become smaller as the overall contingency between the cause and the redundant cue increases proportional to the number of Phase 2 trials.

Method

Participants and design. Thirty students from the University of Munich, Germany, participated in this experiment. They were randomly assigned to either the no-information condition (15 participants) or the explicit-absence condition (15 participants).

Procedure and materials. Participants were run individually on a microcomputer. Prior to the learning task they received written instructions (in German). All participants read diagnostic cover stories in which the cues were characterized as potential effects of a common cause. The instructions stated that the task was to learn about a new disease of the blood, Midosis, which is caused by a virus. This disease was studied in animals. Scientists hypothesized that the disease may cause specific novel substances in the blood. Thus, these substances were characterized as potential effects of the disease. Furthermore it was pointed out that the task involved learning to diagnose the disease on the basis of information about the presence or absence of the substances.

After these initial instructions participants had to give a summary of the instructions, and then the learning phases started. First participants read further instructions on a computer screen in which they were informed that they were going to receive information about individual animals. In the no-information condition it was pointed out that they were going to see information about the presence or absence of two substances, Substance 1 and Substance 2, but that there may be other substances for which no diagnostic tests were conducted.

In Phase 1, all participants received 20 cases of two patterns in random order. Whenever Substance 1 was present and Substance 2 was absent ("Substance 1: Yes; Substance 2: No"), participants had to learn to hit the M key (Midosis); whenever both Substance 1 and Substance 2 were absent ("Substance 1: No; Substance 2: No"), participants were supposed to hit a key that represented the "no" response. After each decision they received corrective feedback about the presence or absence of Midosis. Thus, Substance 1 (predictive cue) was established as a perfect predictor of Midosis. Substance 2 was irrelevant (uncorrelated cue) as it was constantly absent regardless of whether the cause was present or absent. (This cue was added to give participants the opportunity to use the full range of the predictiveness rating scale, which makes it more plausible in the no-information condition that such a scale was to be used.)

Before Phase 2 started, additional instructions were provided

that mentioned that the participants were going to see the results of tests for three substances, two of which were identical with the previous ones. Phase 2 was subdivided into two parts. In each part participants saw 20 cases with information about three substances. With respect to Substances 1 and 2 the trials were identical as in Phase 1. The only difference was that now Substance 3 (redundant cue) was present whenever Substance 1 also was present, and otherwise absent. After Phase 1 and after each segment of Phase 2, ratings of each individual substance were requested. Therefore, participants rated two substances after Phase 1 and three substances after each segment of Phase 2. Similar to Experiment 1, participants were asked to use a rating scale ranging from 0 (*cannot say whether the disease is present*) to 100 (*perfectly certain that the disease is present*) to express how well the substance individually predicts Midosis. No reference to the causal status of the cues and outcomes was made in the rating instructions. However, given that the symptoms were described as effects in the initial instructions the ratings were directed from effects to causes within the causal model participants were expected to form (diagnostic direction).

The explicit-absence condition was similar. The only difference was that the initial instructions mentioned three substances from the outset. Again the possibility of the presence of other substances was mentioned. Phase 1 presented the same trials as the no-information condition except for the fact that additionally Substance 3, the redundant cue in Phase 2, was already mentioned. This substance was described as being absent in each trial of this phase. Phase 2 trials were identical to the ones used in the no-information condition. In the explicit-absence condition participants rated three substances after each learning phase.

Results and Discussion

Figure 2 displays the results of this experiment. As in the previous experiment the predictive cue yielded ratings close to the maximum at all measurement points in both conditions. By contrast, the constantly absent uncorrelated cue was generally given ratings that were close to 0 in both conditions. The most interesting contrast involved the comparison between the predictive and the redundant cue in Phase 2. A 2 (no-information vs. explicit-absence condition) × 2 (predictive cue vs. redundant cue) × 2 (Phase 2a vs. Phase 2b) analysis of variance with the latter two factors being within-subjects factors yielded a reliable three-way interaction, $F(1, 28) = 6.16$, $p < .025$, $MSE = 40.1$. No reliable difference was obtained between the predictive and the redundant cue in the no-information condition. Thus, once again the absence of a blocking effect after diagnostic learning of a common-cause structure with diagnostic test questions was demonstrated. In contrast, the explicit-absence condition in which participants learned about the constant absence of the redundant cue within Phase 1 led to a reduction of the ratings of this cue in Phase 2 although this cue was perfectly correlated with the effect within this learning phase. Within Phase 2a, the first half of Phase 2, the 2 (no-information vs. explicit-absence condition) × 2 (predictive cue vs. redundant cue) interaction proved reliable, $F(1, 28) = 6.11$, $p < .025$, $MSE = 250.4$.

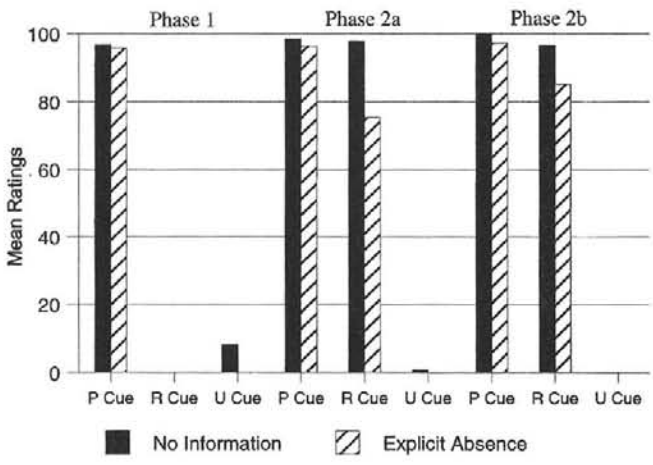

Figure 2. Mean predictiveness ratings for the predictive cue (P cue), redundant cue (R cue), and the uncorrelated cue (U cue) in Experiment 2.

Although, in the explicit-absence condition, the ratings for the redundant cue in Phase 2a were clearly higher than in Phase 1, they were still reliably lower than the ones for the predictive cue, $F(1, 14) = 6.56$, $p < .025$, $MSE = 497.5$. Apparently participants integrated the first two learning phases, which led to a lower contingency estimate for the redundant cue than for the predictive cue after Phase 2a. As predicted by contingency analysis, after Phase 2b (the second half of Phase 2) this difference has further decreased, which reflects the fact that overall the unconditional contingency between the outcome and the redundant cue has increased due to the additional Phase 2 trials. Thus, after this phase the interaction between learning condition and the blocking effect was not significant ($p > .20$).

These results provide clear further evidence of the absence of blocking in the standard diagnostic-learning task with diagnostic test questions, a finding that has been questioned by some researchers (e.g., Matute et al., 1996). Furthermore, this experiment provides evidence for the assumption that the complete absence of blocking in diagnostic conditions is partly based on retrospective inferences about the relation between the cause and the redundant effect in earlier learning phases. This inference is implied by the structure of common-cause models and by assumptions about the temporal stability of causal power, and therefore is a side effect of sensitivity to causal directionality. Such retrospective inferences about the presence of the redundant cue in Phase 1 require cover stories and trial descriptions that clearly distinguish between explicit information about presence or absence and noninformation about the status of an event. Finally, the results of the experiment once again weaken the alternative interpretation that failures to obtain blocking in the diagnostic condition may be due to the possibility that participants treat the two phases in blocking paradigms as unrelated (Shanks & Lopez, 1996). The explicit-absence condition clearly shows that participants were affected by the trials in Phase 1 when they gave ratings in Phase 2. The absence of blocking in diagnostic learning can therefore not be attributed to particular features of two-phase blocking tasks.

Many associative theories do not differentiate between the absence of cues and the absence of information about cues (e.g., Rescorla & Wagner, 1972), and therefore do not

predict a difference between the no-information and the explicit-absence conditions. Some more recent theories try to incorporate this difference, for example by coding absence with a −1 or some other negative number and noninformation with a 0, or by postulating different learning rates (see Tassoni, 1995; Van Hamme & Wasserman, 1994). Dickinson and Burke (1996) suggested that the omission of a cue that is expected because it has been previously seen as part of a compound causes retrospective revaluations of the previous status of this cue. However, none of these theories predicts that compound training in a later learning stage leads to the retrospective inference that an element of the compound was already present in a previous learning stage in which it was not shown. Thus, none of these theories predicts the absence of blocking in the no-information condition of diagnostic blocking paradigms.

Experiment 3a

To further differentiate between predictions of associative learning theories and causal-model theory, Experiment 3 introduces a new blocking design in which a redundant cue competes with either one or two initially learned causal relations (see also Waldmann, 1996, for a short summary of Experiment 3a). This new paradigm allows it to directly test the assumption of causal-model theory that diagnostic ratings are sensitive to the presence of alternative explanations of the effect. Diagnostic test questions should only yield equal ratings of the predictive and redundant cue within blocking designs when the contingencies between the common cause and each effect are equated, and when the effects cannot differentially be explained by alternative causes. Waldmann and Holyoak (1992) only indirectly supported the latter assumption on the basis of a cross-experiment comparison. Furthermore, the predictive-learning conditions of this experiment permit a test of additional implications of causal-model theory.

Table 1 (top) displays the task in the diagnostic conditions. The instructions and materials were similar to the ones used in Experiment 2. In the learning phases, participants received information about the presence or absence of different substances in animals' blood. Participants' task was to learn whether or not the animal had contracted one of two new blood diseases.

In the diagnostic conditions, the substances were described as effects of the diseases. Participants were told that new blood diseases had been discovered that produce new types of substances in the blood. In both conditions, the unambiguous and the ambiguous cue condition, participants learned in Phase 1 that Substance 1 is caused by Disease 1, and Substance 2 is caused by Disease 2. In Phase 2, however, the two conditions were different. In the unambiguous cue condition, Substance 3 is redundantly paired only with Substance 1. Participants learned that Disease 1 apparently has two effects, Substance 1 and Substance 3.

In the test phase, participants were requested to rate how predictive each individual substance is for each disease. Associative learning theories, such as the Rescorla–Wagner theory, predict blocking (i.e., different ratings of Substance 1 and Substance 3) in the unambiguous cue condition (also Matute et al., 1996; Price & Yates, 1995; Shanks & Lopez, 1996). By contrast, causal-model theory predicts absence of blocking because both substances are deterministic effects of the disease and because there are no alternative competing explanations for the presence of Substance 3.

In the ambiguous cue condition, Substance 3 is caused by either Disease 1 or Disease 2. Associative theories again predict complete blocking of the redundant cue. By contrast, causal-model theory implies that participants should be sensitive to the fact that there are multiple explanations for the presence of Substance 3. Therefore it is expected that participants will lower their diagnostic ratings in this condition. However, this lowering is, according to causal-model theory, not due to competition between the redundant effect, Substance 3, and the predictive effects, Substances 1 and 2; it is a consequence of the fact that Substance 3 is potentially produced by two mutually exclusive causes. This fact limits its diagnostic value.

A predictive version of the task was also investigated to provide further evidence for participants' sensitivity to causal directionality. Table 1 (bottom) outlines the structure of the task in the predictive conditions. In these conditions, the very same cues and outcomes were used as in the diagnostic conditions, only the causal direction connecting cues and outcomes was reversed in the cover story. In these conditions the substances were redefined as potential causes of the new blood diseases. Participants were instructed that some food items appear to contain substances that may cause new blood diseases. The same learning exemplars were used as in the diagnostic conditions. Thus, in Phase 1, participants learned that Substance 1 causes Disease 1, and Substance 2 causes Disease 2. Again in Phase 2 two different

Table 1
Design of Experiment 3a

Phase	Cue condition: Unambiguous	Cue condition: Ambiguous
	Diagnostic learning	
1	$Effect_1 \leftarrow Cause_1$ $Effect_2 \leftarrow Cause_2$	$Effect_1 \leftarrow Cause_1$ $Effect_2 \leftarrow Cause_2$
2	$Effect_1 + Effect_3 \leftarrow Cause_1$ $Effect_2 \leftarrow Cause_2$	$Effect_1 + Effect_3 \leftarrow Cause_1$ $Effect_2 + Effect_3 \leftarrow Cause_2$
	Predictive learning	
1	$Cause_1 \rightarrow Effect_1$ $Cause_2 \rightarrow Effect_2$	$Cause_1 \rightarrow Effect_1$ $Cause_2 \rightarrow Effect_2$
2	$Cause_1 + Cause_3 \rightarrow Effect_1$ $Cause_2 \rightarrow Effect_2$	$Cause_1 + Cause_3 \rightarrow Effect_1$ $Cause_2 + Cause_3 \rightarrow Effect_2$

Note. The same cues (substances) were used as effects in the diagnostic-learning condition and as causes in the predictive-learning condition. Also the outcomes (causes or effects) in both conditions were identical (diseases). The different causal roles were manipulated solely by means of initial instructions. The arrows represent the causal arrow, which is always directed from causes to effects. Learning order was constant. Events on the left side of the arrows (i.e., effect cues in diagnostic learning or cause cues in predictive learning) uniformly represent the cues that were presented first; the events described on the right side of the arrow represent the outcomes that were shown after participants' responses.

structures were compared: In the unambiguous cue condition Substance 3 was either redundantly paired only with Substance 1 to produce Disease 1; or, in the ambiguous cue condition, it was paired with either Substance 1 to produce Disease 1 or Substance 2 to produce Disease 2. Also the same rating instructions were used as in the diagnostic conditions.

Assuming that Phase 1 training reached the learning asymptote, associative learning theories predict that the redundant cue should not gain any associative weight in either the ambiguous or the unambiguous cue condition. Blocking should be complete and equal in both conditions. Causal-model theory predicts that in both conditions the cues will be assigned the causal status of potential causes. Some of these potential causes potentially converge on a common effect (e.g., Substances 1 and 3). This should caution participants to take possible interactions between causes into consideration. In the unambiguous cue condition, the unconditional contingency between Substance 3 and Disease 1 is high (1.0), which suggests this substance as a potential cause. However, participants always see the redundant cue together with a previously established deterministic cause, which makes it impossible to compute the conditional contingency between the redundant cause and the effect in the absence of the initially predictive cause. Furthermore, the predictive cues deterministically cause the effects, which makes it impossible to observe the potential influence of redundantly paired causes. Thus, it is impossible to determine whether the redundant cue represents a spurious cause or whether this is a situation of causal interaction or overdetermination. In this situation participants should be reluctant to accept the redundant cue as individually predictive. They are expected to give ratings that relative to the predictive cue express their uncertainty about the causal status of the redundant cue (i.e., partial blocking).

In the ambiguous cue condition, the unconditional contingency between the redundant cue and Disease 1 is lower (0.5). Also, tests of conditional contingencies yield different outcomes in the two conditions. Whereas in the unambiguous cue condition, no information is presented about the conditional probability of Disease 1 given Substance 3 in the absence of the predictive Substance 1, in the ambiguous cue condition participants learn that this conditional probability is 0. In this condition, participants actually see Substance 3 in the absence of either predictive substance whereas in the unambiguous condition, Substance 3 never is present when the corresponding predictive substance is absent. The information that the effect does not occur when the redundant cue is present without the corresponding predictive cue is predicted to be taken as evidence by participants that the redundant cue is probably not an individual cause of the diseases. This should lead to relatively lower ratings of the redundant cue in the ambiguous than the unambiguous cue condition.

The present experiment deviates from the standard blocking paradigm as no control condition was used that only presented Phase 2 trials. As pointed out by Chapman and Robbins (1990) the comparison between the blocking condition and this possible control condition is confounded by an unequal number of observations of the outcome in the absence of the redundant cue. Such observations are presented in Phase 1 of the blocking condition but not in the suggested control condition. Therefore, Chapman and Robbins (1990) proposed the control that, slightly adapted, was also used in the present Experiment 1. This experiment clearly demonstrated blocking in the predictive-learning condition, and absence of blocking in the diagnostic-learning condition. In Experiment 3a a different type of test of causal-model theory was used. Two learning conditions (predictive vs. diagnostic) were compared that presented identical trials and identical test questions. Associative theories predict identical learning whereas causal-model theory predicts an interaction between the blocking effect and the causal learning conditions. Furthermore, causal-model theory predicts the absence of a blocking effect in the unambiguous-cue condition, which is incompatible with the associationist prediction of lower ratings for the cue that has been seen fewer times.

Method

Participants and design. Participants were 56 students from the University of Tübingen, Germany, who were randomly assigned to the four conditions outlined in Table 1 (14 participants per condition).

Procedure and materials. Participants were run individually on a microcomputer. Prior to the learning task they received written instructions (in German). Participants in the diagnostic conditions were told that they were going to learn about two new diseases of the blood, Midosis and Xeritis, which are caused by viruses. These diseases were studied in animals. Scientists hypothesized that the diseases may cause specific new types of substances in the blood. Thus, these substances were characterized as potential effects of the diseases. The predictive learning instructions were similar. According to this cover story, scientists hypothesized that different new types of substances, which occur in specific food items and have been detected in the blood of sick animals, may be the causes of the diseases. Therefore, the scientists fed animals with different combinations of the substances in order to see whether they would contract any of the new diseases. In this condition, the substances played the role of potential causes of the diseases.

The learning phases were identical in the predictive- and diagnostic-learning conditions. In Phase 1, all participants received 30 cases of three patterns in random order. Whenever Substance 1 was present and Substance 2 was absent ("Substance 1: Yes; Substance 2: No"), participants had to learn to hit the M key (Midosis); whenever Substance 2 was present while Substance 1 was absent ("Substance 1: No; Substance 2: Yes") participants were supposed to hit the X key (Xeritis). When both substances were absent ("Substance 1: No; Substance 2: No"), the space bar ("none of the diseases") was the appropriate response. No other cases were presented within this learning phase. After each decision participants received corrective feedback.

After Phase 1, participants were handed rating instructions that were identically phrased in both the predictive- and diagnostic-learning conditions. The task was to assess on a rating scale ranging from 0 (*cannot say whether the disease is present*) to 100 (*perfectly certain that the disease is present*) how well the presence of each of the substances individually predicts Midosis or Xeritis.

Phase 2 was divided into two blocks. In each block, participants saw seven cases of Midosis, seven cases of Xeritis, and seven

no-disease cases in random order. The structure of the items with respect to Substances 1 and 2 was the same as in Phase 1. These two substances still were perfect predictors of the diseases. The only difference was that a new substance, Substance 3, was either redundantly present when either of the previous substances was present (ambiguous cue condition), or it was only present when one of the previous substances was present (unambiguous cue condition). The substance with which Substance 3 co-occurred in the unambiguous cue condition was counterbalanced across participants. Thus, in the unambiguous cue condition, one of the subconditions presented the following three cases (A, B, C) as learning trials in Phase 2:

A. Substance 1: Yes; Substance 2: No, Substance 3: Yes → Midosis,
B. Substance 1: No; Substance 2: Yes; Substance 3: No → Xeritis,
C. Substance 1: No; Substance 2: No; Substance 3: No → no disease.

For the ambiguous cue condition, only Case B was altered:

Substance 1: No; Substance 2: Yes; Substance 3: Yes → Xeritis.

It is notable that in the presentation of the cues explicit mention of the absence of a substance (e.g., "Substance 1: No") is clearly distinguishable from noninformation about a substance (see Experiment 2). After each block, participants were requested to rate the predictiveness of each of the three substances individually with respect to each disease.

Results and Discussion

Figure 3 shows that on average the predictive cue (i.e., the cue or the cues with which the redundant cue was later paired) received high mean ratings in the four conditions, demonstrating that participants were able to learn the predictive relations between the two predictive substances and diseases. These ratings were statistically equivalent ($F < 1$). Figure 3 also depicts the mean ratings of the different substances in the diagnostic- and the predictive-learning conditions that were obtained in Phase 2. Because the two measurements within Phase 2 did not yield any reliable differences, the ratings were averaged across the two measurements. The redundant cue represents the mean ratings of Substance 3 with respect to one (unambiguous cue condition) or both diseases (ambiguous cue condition) with which it co-occurred. A 2 (predictive vs. diagnostic learning) × 2 (ambiguous vs. unambiguous cue condition) × 2 (predictive vs. redundant cue) analysis of variance with the latter factor referring to a within-subjects manipulation yielded a marginally significant three-way interaction, $F(1, 52) = 3.52$, $p < .07$, $MSE = 335.4$, and a significant two-way interaction between the blocking factor (predictive vs. redundant cue) and the structure of the learning input (ambiguous vs. unambiguous), $F(1, 52) = 49.2$, $p < .001$, $MSE = 335.4$.

As can be seen in Figure 3, no attenuations of the ratings occurred for the redundant cue relative to the predictive cue within the diagnostic unambiguous cue condition. There was no sign of blocking: In fact, as in Experiment 1, all 14 participants gave identical ratings for the predictive and the redundant cue in this condition. As predicted, participants learned that Substance 1 as well as Substance 3 is a reliable diagnostic indicator of the disease. The lack of a blocking effect refutes the prediction of standard associative theories, such as the Rescorla–Wagner theory. Using a different blocking paradigm, this result along with the results of

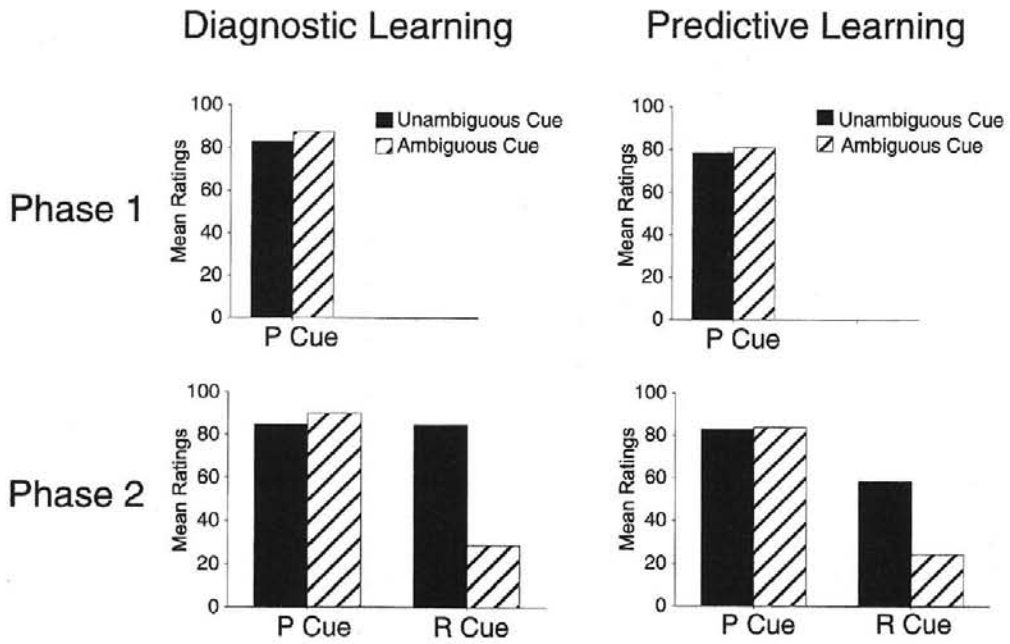

Figure 3. Mean predictiveness ratings (averaged over two measurements) for diagnostic and predictive conditions in Phases 1 and 2 of Experiment 3a, for the predictive cue (P cue) and the redundant cue (R cue).

Experiments 1 and 2 clearly replicates the finding in Experiment 3 of Waldmann and Holyoak (1992), absence of blocking after diagnostic learning with diagnostic test questions, that was questioned in a number of critical responses.

Furthermore, the results show that participants in the diagnostic ambiguous cue condition lowered their ratings for the redundant cue. An analysis of variance in which the predictive and the redundant cues were compared within the diagnostic ambiguous cue condition yielded a significant effect, $F(1, 13) = 73.5, p < .001, MSE = 360.6$. The ratings of the redundant cue were clearly lower in the ambiguous cue condition relative to the unambiguous cue condition, $F(1, 26) = 34.7, p < .001, MSE = 634.4$. The reduced ratings for the ambiguous redundant cue relative to the rating of the same cue in the unambiguous cue condition signify that participants did not simply express the cause-to-effect contingency in their ratings (which is 1 in both conditions). Apparently, they realized that the diagnostic validity of a cue is dependent on the presence of alternative possible explanations of the presence of the cue. Since the judged substance is potentially caused by either disease it is normatively correct to lower the predictiveness rating of the redundant cue with respect to either disease. The difference between the diagnostic ambiguous and the unambiguous cue condition provides direct evidence for causal-model theory's assumption that participants take the possibility of alternative causes into account when making diagnostic judgments. The different patterns in the two diagnostic conditions also refute variants of associative learning theories that suggest a general mapping of causes to the input level (Van Hamme et al., 1993). Mapping the causes of both the ambiguous and the unambiguous cue condition to the input level would predict equal ratings (i.e., absence of cue competition) in both conditions.

The pattern of results differed in important ways for the predictive-learning conditions. In general, there was a clear attenuation of the ratings for the redundant cue in the predictive-learning conditions. A 2 (predictive vs. redundant cue) × 2 (ambiguous vs. unambiguous cue condition) analysis of variance with the first factor referring to a within-subjects manipulation yielded a reliable difference between the ratings of the predictive and redundant cues, $F(1, 26) = 49.2, p < .001, MSE = 490.6$, as well as a significant interaction between the two factors, $F(1, 26) = 9.03, p < .01, MSE = 490.6$. A comparison of the ratings of the predictive and redundant cue within the unambiguous predictive-learning condition also proved to be statistically significant, $F(1, 13) = 9.79, p < .01, MSE = 403.5$. Furthermore, the redundant cue was rated significantly lower in the unambiguous predictive-learning condition than in the unambiguous diagnostic-learning condition, $F(1, 26) = 5.39, p < .05, MSE = 899.7$, whereas no reliable difference was found in the corresponding ambiguous conditions ($F < 1$). Finally, an analysis of variance in which the redundant cue was compared across the ambiguous and unambiguous cue conditions of the predictive-learning condition also turned out to be significant, $F(1, 26) = 10.3, p < .01, MSE = 781.8$. Thus, as predicted by causal-model theory as well as associative learning theories, ratings for the redundant cue were attenuated in both conditions. However, causal-model theory additionally predicts the difference of the ratings for the redundant cue between the ambiguous and the unambiguous cue condition.

Associative theories could explain the results in the predictive condition, however, if the assumption was made that learning in Phase 1 was in fact pre-asymptotic. It is unlikely that this account is correct. Participants only had to learn to associate three simple patterns with three responses. This is an extremely easy task that was mastered by most participants within a couple of trials. Furthermore, this account would predict an increase of the ratings of the predictive cue with increasing training, which was not observed (see also Waldmann & Holyoak, 1992). Nevertheless, Experiment 3b will provide an empirical test of this alternative explanation of the results of the predictive-learning conditions.

The reliable interaction between the blocking effect and causal learning condition also takes care of other counterarguments against two-phase blocking designs. Shanks and Lopez (1996) argued that the blocked presentation in two-phase designs may lead participants to treat the two phases as unrelated. Williams et al. (1994) also pointed out the problem that blocking sometimes may not be obtained due to a configuration of cues presented within Phase 2. In a similar vein, one anonymous reviewer proposed the hypothesis that the unambiguous cue condition may elicit configural processing (i.e., absence of blocking), whereas the ambiguous cue condition may elicit elemental processing (i.e., blocking) which also could explain the differences in the size of the blocking effect in these two conditions. However, all these alternative associationist accounts are unable to account for the basic finding that with identical learning materials the blocking effect interacted with the causal condition. All these alternative accounts either predict uniform results or a main effect across the ambiguous and unambiguous cue conditions but not the interaction with the causal framing of the task that was actually observed. Only causal-model theory is consistent with the full pattern of results.

Experiment 3b

Experiment 3b focuses on the blocking effect in the predictive-learning condition. On the basis of contingency analyses, causal-model theory predicts blocking in this condition as a consequence of the unavailability of relevant information for assessing the causal status of the redundant cue. The Rescorla–Wagner rule and related associative learning theories consider blocking a result of the fact that strong valid cues tend to suppress learning of redundant cues. Empirically these two theories may be distinguished by looking at the amount of blocking. Causal-model theory predicts partial blocking as participants are expected to be uncertain about the causal status of the redundant cue rather than be certain that it is not a cue. The Rescorla–Wagner rule predicts complete blocking of the redundant cue as a result of the predictive cue having reached the asymptote in Phase 1.

A further test to distinguish between these two accounts was provided by Experiment 3a. In this experiment a condition in which a redundant cue was potentially blocked by two predictive cues (ambiguous cue condition) was compared with the standard case in which the redundant cue was potentially blocked only by one cue (unambiguous cue condition). Again the Rescorla–Wagner theory (and many other associative theories) predict full blocking in both conditions, whereas causal-model theory predicts partial blocking with lower ratings in the ambiguous relative to the unambiguous cue condition. The results of Experiment 3a clearly favored causal-model theory.

However, it could be argued that the Rescorla–Wagner rule also predicts this pattern in the predictive-learning condition if it is assumed that the predictive cues were not trained to the asymptote within Phase 1. Rescorla and Wagner's (1972) theory views blocking as the result of a failure to acquire associative strength. According to this rule learning is error driven. The Rescorla–Wagner rule

$$\Delta V_i = \alpha_i \beta_j (\lambda_j - \Sigma V)$$

states that the change in associative strength on a trial for each presented cue i, ΔV_i, is proportional to the difference between the outcome concerning US_j that should have been predicted, λ_j, and that predicted by the sum of all current cues, ΣV, weighted by learning rate parameters α_i and β_j that are specific to the particular CS and the US, respectively. In blocking experiments animals learn to predict the outcome with the initially acquired predictive cue CS_1. Since this cue still enables perfect predictions in Phase 2 no further learning occurs. The CS_2 stays at its initial value. However, only when the associative weight for the predictive cue V_P is trained to the maximum weight λ, no further change is expected for the predictive and the redundant cues during Phase 2. If V_P is smaller than λ after Phase 1, then the error term at the beginning of Phase 2, $(1 - [V_P + V_R])$, still yields a positive value (assuming that V_R, the associative weight of the redundant cue, initially is at a weight of 0). Thus, both weights will grow until their sum reaches the learning asymptote.

Furthermore, the Rescorla–Wagner theory predicts a difference between the ambiguous and the unambiguous cue conditions when the predictive cue is preasymptotic after Phase 1 (see Appendix for mathematical derivations). In the unambiguous cue condition the maximal weights for the predictive cue and the redundant cue after Phase 2, $V_{P\text{-max}}$ and $V_{R\text{-max}}$, are a function of the difference of the learning asymptote λ and the weight of the P cue after Phase 1, $V_{P\text{-Phase1}}$. Assuming λ to be 1 and the learning rates to be equal, then the predictive and the redundant cue divide the maximum weight left after Phase 1 among each other,

$$\Delta V_P = \Delta V_R = \frac{(1 - V_{P_{Phase1}})}{2}.$$

In the ambiguous cue condition the maximal weight gains for the predictive and the redundant cue are also determined by the distance of the weight of the predictive cue from the learning asymptote at the end of Phase 1. However, in this condition the redundant cue not only grows when it is paired with one of the outcomes in Phase 2, it also decreases as a function of being paired with the alternative outcome. For example, when this cue is paired with the second outcome, Disease 2, it is simultaneously paired with the absence of the first outcome, Disease 1. Thus, the associative weight that links this cue to Disease 1 is decreased after this trial. No such decreases occur for the predictive cue. Thus, assuming equal learning rates the predictive cue will gain weights twice as fast as the redundant cue (see Appendix),

$$\Delta V_P = \frac{2(1 - V_{P_{Phase1}})}{3}; \Delta V_R = \frac{(1 - V_{P_{Phase1}})}{3}.$$

Therefore, assuming preasymptotic learning in Phase 1, the associative weight of the predictive cue should rise faster in the ambiguous cue condition than in the unambiguous cue condition, whereas the weight for the redundant cue should rise more slowly in the ambiguous cue condition. Furthermore, the differences between these two conditions should be more pronounced the larger the difference between the weight for the predictive cue and the learning asymptote turns out to be before Phase 2 training commences.

These predictions are tested in Experiment 3b in which a predictive-learning task and the causal structures from the unambiguous and the ambiguous cue conditions from Experiment 3a were used. The only difference was that an uncorrelated cue was added whose presence and absence was paired with the absence of either disease. In this experiment two learning conditions were compared. In the "short" learning condition participants saw each case only twice within Phase 1; in the "long" condition they received each case 10 times. Thus, if it is assumed that the number of trials used in Experiment 3a yielded preasymptotic learning, a trial number of only two trials should lead to associative weights that are even further away from the asymptote. The choice of the shorter learning phase was based on the assumption that because of the diminishing increase characteristic of the predicted associative learning curve it should be easier to observe differences in the early phases of training than in later phases (e.g., 30 vs. 40 trials), provided that associative theories are correct. The variation of this factor permits the test of the assumption of the Rescorla–Wagner rule that the size of the blocking effect should be dependent on the length of Phase 1 training. In Phase 2 two measurements of the acquired causal strengths were obtained. The first measurement (Phase 2a) was already taken after each case again was presented twice, the final ratings (Phase 2b) were obtained after each case was presented another 10 times.

The predictions of the Rescorla–Wagner rule and causal-model theory differ for these conditions. Provided that the associative weight of the predictive cue is further away from the asymptote after the short Phase 1 training than after the long training regime, the Rescorla–Wagner rule predicts that the associative weights should grow between Phase 2a and

Phase 2b, the two measurement points in Phase 2. However, this increase should be larger in the short condition in which the predictive cue ends up further away from the asymptote after Phase 1. Furthermore, the blocking effect is expected to be smaller in the short than in the long condition. The partial blocking effect in the short condition is a joint function of lower weights for the predictive cue and higher weights for the redundant cue. In the long condition the weights for the redundant cues should end up close to the ones for the uncorrelated cue, which should stay at 0 in all conditions. Finally, the difference between the redundant cue in the ambiguous and the unambiguous cue condition is expected to be larger in the short training condition, and should increase with additional Phase 2 training.

Causal-model theory does not predict these patterns. According to this theory, causal strength is estimated on the basis of conditional and unconditional contingencies. These contingencies do not change as a function of short or long Phase 1 training, although the confidence in the estimates may change as a function of sample size. The partial blocking effect and the difference between the ambiguous and the unambiguous cue condition is predicted on the basis of differential availability of relevant information to assess the causal status of the cues, and is not dependent on the length of Phase 1 training.

Method

Participants and design. Eighty students from the University of Tübingen, Germany, participated in this experiment. They were randomly assigned to one of the four conditions that were generated by crossing the factor of length of Phase 1 (short vs. long) with the factor of causal structure (ambiguous vs. unambiguous cue condition). There were 20 participants in each condition.

Procedure and materials. Instructions and materials corresponded to the ones used in the predictive-learning conditions of Experiment 3a, except that an uncorrelated cue was added. In Phase 1, participants saw four types of cases with three substances as potential causes of the two diseases. Either all substances were absent ("no disease") or only one of the three substances was present. One of the two relevant substances (predictive cues) was paired with the disease Midosis, the other one with the disease Xeritis. The presence of the third substance (uncorrelated cue) was paired with the absence of either disease (as its absence).[3] In Phase 2, a fourth substance (redundant cue) was added that, as in Experiment 3a, was perfectly correlated with only one of the predictive cues (unambiguous cue condition), or it was paired with either predictive cue (ambiguous cue condition). The assignment of the three substances to the roles of the two predictive cues and the uncorrelated cue, and the assignment of the redundant cue to one of the predictive cues in the unambiguous cue condition was counterbalanced. In the "short" condition, Phase 1 consisted of eight trials in which each of the four types of cases was presented twice. In the "long" condition, 40 trials were presented with each case being presented 10 times.

After Phase 1, ratings of the different substances were obtained. Participants were asked to imagine that only one of the three substances was given to an animal. To emphasize that only one of the substances had been given, the rating sheets presented patterns of the three substances similar to those presented in the learning phases. Only one substance was declared to be present for the ratings. The other substances were described as being explicitly absent. The sequence of the test cases was randomized. Using a scale that ranged from −10 (*certainly not*) to +10 (*certainly*), participants had to assess how certain they were that the animal had contracted the respective blood disease.

In Phase 2, two measurements were obtained. The first measurement was taken after two presentations of each case (Phase 2a). Then after an additional 10 presentations of each case the final ratings were collected (Phase 2b). The rating instructions were identical to the ones used in Phase 1 except that the new substance from Phase 2 was also included. Unlike during learning, however, only one substance was described as being present, whereas the others were explicitly declared absent.

Results and Discussion

The results of this experiment are displayed in Figure 4. As in Experiment 3a, the statistical analyses are based on the averages of equivalent measurements of the predictive (i.e., the respective predictive cues for either disease), redundant (i.e., the two ratings of the redundant cue in the ambiguous cue condition), and uncorrelated cues (the two ratings with respect to the two diseases).

High ratings were generally obtained for the predictive cues. No increase of the ratings was observed between the short and the long condition. A 2 (ambiguous vs. unambiguous cue condition) × 2 (long vs. short) × 3 (Phase 1 vs. Phase 2a vs. Phase 2b) analysis of variance with learning phase as a repeated measurement factor yielded no significant effects of the ratings involving the predictive cues. (There is a descriptive tendency of lower ratings for the predictive cue in the ambiguous than the unambiguous cue condition [$p < .06$]. However, most likely this difference occurred by chance as this tendency can already be seen in Phase 1 in which the trials of the ambiguous and the unambiguous cue conditions were identical.) Thus, two presentations of each pattern were sufficient to elicit the impression that the predictive substances were strong predictors of the diseases.

To test the impact of the length of Phase 1 on the blocking effect, a 2 (unambiguous vs. ambiguous cue condition) × 2 (long vs. short) × 2 (predictive cue vs. redundant cue) × 2 (Phase 2a vs. Phase 2b) analysis of variance of the ratings was conducted with the last two factors constituting within-subjects variations. The results were clear: The length of Phase 1 training did not have an effect on the obtained patterns. Neither the main effects nor the interactions involving this factor proved reliable. The same holds true for the differences between Phase 2a and Phase 2b. Overall the blocking effect (i.e., the difference between the predictive and the redundant cue) turned out to be highly significant, $F(1, 76) = 521.9$, $p < .001$, $MSE = 27.0$. The effect of causal structure (ambiguous vs. unambiguous cue condition) also proved highly reliable, $F(1, 76) = 17.0$, $p < .001$,

[3] It can be argued that the uncorrelated cue is actually negatively correlated as the disease is sometimes present in the absence of this cue. However, in a situation in which the cues represent causes the best indicator of causal strength is the contingency of a cause in the absence of alternative causes (Cheng, 1997). In the absence of the causes of the two diseases (i.e., the other two substances) the "uncorrelated" cue clearly has a zero contingency. Similarly, the Rescorla–Wagner rule would assign this cue a weight of zero as it does not update weights in the absence of a cue.

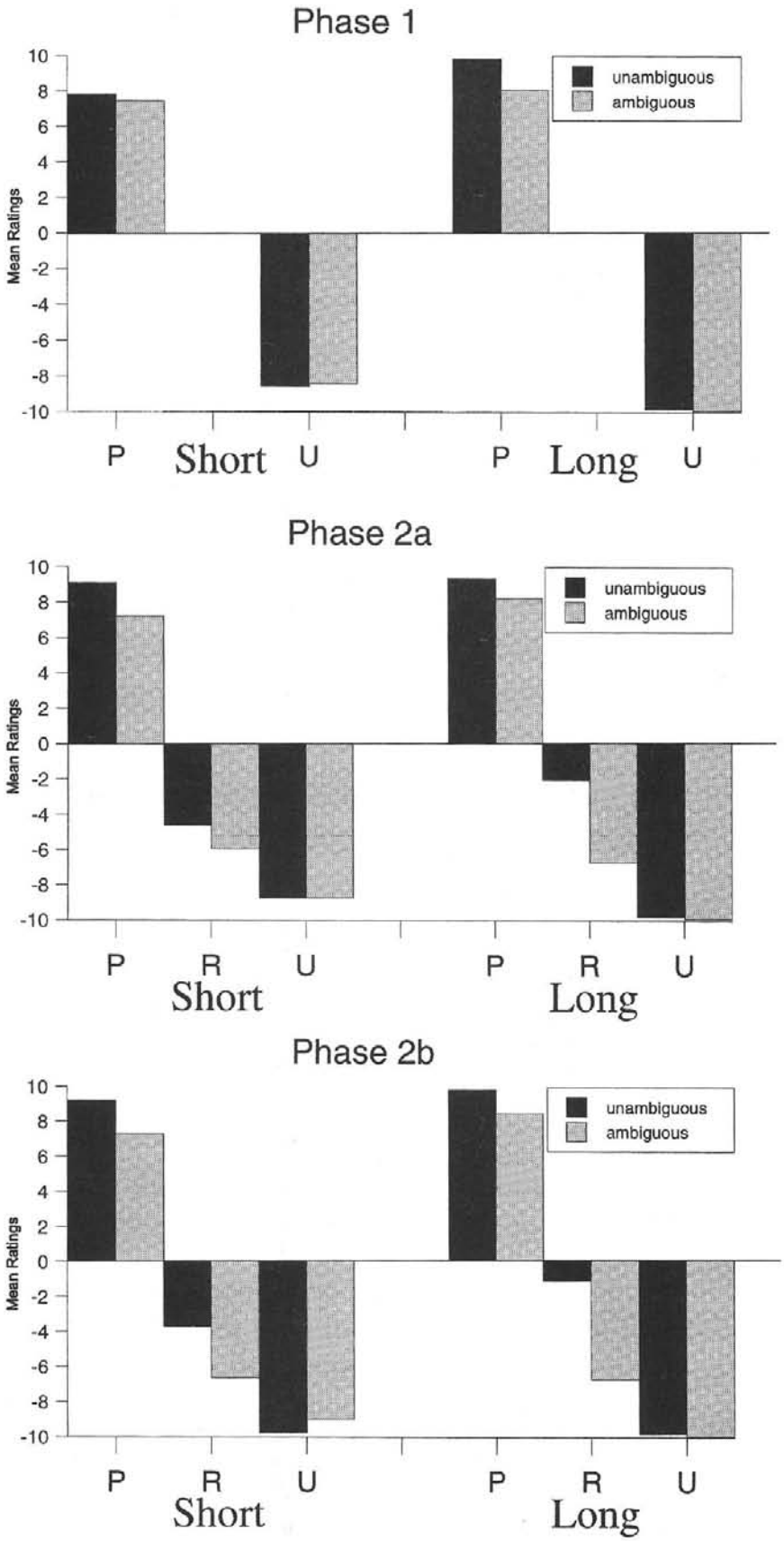

Figure 4. Mean predictiveness ratings for predictive-learning task (Experiment 3b) in Phases 1, 2a, and 2b after short and long Phase 1 training, for the predictive cue (P), the redundant cue (R), and the uncorrelated cue (U).

MSE = 32.4, which is moderated by a marginally reliable interaction between causal structure and blocking, $F(1, 76) = 3.42$, $p < .07$, MSE = 27.0. Again the ambiguous redundant cue was on average rated reliably lower than the unambiguous redundant cue, $F(1, 78) = 14.3$, $p < .001$, $MSE = 19.1$.[4]

The comparison between the uncorrelated cue and the redundant cue constitutes a further test between the competing theories. Causal-model theory predicts partial blocking, that is, a difference between the redundant cue and the uncorrelated cue, independent of length of Phase 1, whereas the Rescorla–Wagner theory expects a decrease of the difference between the redundant and the uncorrelated cue proportional to the number of Phase 1 trials. The results show that the redundant cue was generally rated higher than the uncorrelated cue, $F(1, 76) = 96.2$, $p < .001$, MSE = 19.0. Due to the pronounced difference between the ratings of the redundant cue in the unambiguous and the ambiguous cue conditions the interaction between the ambiguity factor and the cue factor (redundant vs. uncorrelated cue) also was reliable, $F(1, 76) = 15.4$, $p < .001$, MSE = 19.0. In this analysis, the length of Phase 1 additionally interacted with the size of the difference between the redundant and the uncorrelated cue, $F(1, 76) = 4.01$, $p < .05$, MSE = 19.0. This result, however, is a consequence of the fact that the difference between the averaged redundant cues and the uncorrelated cue is slightly larger in the long Phase 1 condition than in the short condition. It is important to note that this tendency of a decrease of the size of the blocking effect with increasing Phase 1 training is opposite to the pattern the Rescorla–Wagner theory would predict.

The results of this experiment favor causal-model theory over associative accounts. Associative theories make the counterintuitive prediction that few trials indicating a deterministic causal relation should yield the same ratings as many trials that are based on a probabilistic relation. Contrary to this view, the results of this experiment show that a few trials were sufficient to create the impression that the relevant substances were strong causal predictors of the diseases. Also this experiment provides direct evidence against the alternative associationist theory that partial blocking and the difference between the ambiguous and the unambiguous cue conditions in Experiment 3a were due to preasymptotic learning of the predictive cues within Phase 1. The variation of the length of Phase 1 did not have any statistically reliable effects on the difference between the predictive and the redundant cue. Furthermore, if anything, the descriptive results seem to be opposite to the pattern predicted by the Rescorla–Wagner rule. The difference between the ratings of the ambiguous and the unambiguous redundant cues seems to be slightly larger in the condition in which Phase 1 was longer (see Figure 4). The Rescorla–Wagner rule, however, predicts that the difference between these ratings should be smaller in this condition. One possible explanation for this descriptive pattern is that participants may have been less sure about the causal status of the cues after brief training. After all, the ambiguous and the unambiguous cue conditions differ only with respect to one case. The more training exemplars participants saw, the easier it may have become to differentiate between the cues. This is only one possible suggestive interpretation of the observed patterns.

A possible alternative associationist explanation of the results could claim that learning proceeded so fast that the asymptote had already been reached after the eight learning trials of Phase 1 in the "short" condition. However, this explanation is undercut by the observed reliable difference of the ratings of the redundant cue in the ambiguous versus the unambiguous cue conditions observed in both Experiments 3a and 3b. According to the Rescorla–Wagner rule blocking should be complete in both conditions after asymptotic Phase 1 learning. Thus, associationist theories either may claim that fast learning precluded the observation of differences between the "short" and "long" conditions, which would entail the prediction of full blocking in both the ambiguous and unambiguous cue conditions, or the observed differences might be explained by invoking preasymptotic learning, an account that is weakened by the fact that training length in Phase 1 did not have an effect on Phase 2 learning.

General Discussion

Causal asymmetry is one of the most fundamental features of the physical world. Causes are the interface of our interactions with physical reality, whereas effects can only indirectly be manipulated. Furthermore, multiple causes of a common effect may interact or collaborate in producing an effect, whereas multiple effects of a common cause are typically conditionally independent unless influenced by additional hidden factors. Thus, the ability to learn about these fundamental causal relations correctly is central to arrive at adequate mental representations of our world. This article is part of a debate between the view that we are sensitive to causal directionality during learning (causal-model theory) and the view that our learning only involves acquiring associations between cues and outcomes irrespective of their causal role (associative theories). Causal-model theory claims that we have the competence to represent this aspect of physical reality correctly, whereas associative theories predict that, at least in some learning conditions, our representations are fundamentally flawed.

Waldmann and Holyoak (1992) presented evidence in favor of causal-model theory. However, their theoretical conclusions and the validity of their experimental results have been questioned. The goal of the present article is to provide further evidence bearing on the debate between

[4] In this experiment some of the cues yielded negative ratings, which raises the question whether participants thought that these cues were involved in negative contingencies. However, the rating instructions only focused on the case that the individual substances were present. Thus, a negative rating only expressed certainty that the disease is absent in the presence of the substance, no assumptions about the status of the disease in the absence of the substance are implied (which is necessary for expressing negative contingencies in this situation). Thus, negative ratings correspond to low ratings (near zero) in the scales used in the previous experiments.

causal-model theory and associative accounts, and to elaborate and test additional implications of causal-model theory.

Summary and Discussion of Results

A number of critics have questioned the validity of the most important result of Waldmann and Holyoak (1992): the absence of a blocking effect after diagnostic learning and diagnostic testing (Matute et al., 1996; Price & Yates, 1995; Shanks & Lopez, 1996). Although the absence of blocking with effects has been acknowledged when the effects are presented as outcomes, most associative accounts predict blocking when the effects are presented as cues and the test questions are also directed from cues to outcomes (i.e., from effects to causes). In Experiment 1 (diagnostic condition), Experiment 2 (no-information condition), and Experiment 3a (diagnostic unambiguous cue condition), which represent such conditions, the predicted absence of blocking was clearly replicated.

Sometimes the small descriptive difference found in the diagnostic condition of Experiment 3 of Waldmann and Holyoak (1992) was interpreted as evidence for blocking (Matute et al., 1996; Shanks & Lopez, 1996). Experiment 2 offers an alternative account based on an elaboration of causal-model theory's predictions. Causal-model theory claims that in common-cause situations assessments of the relation between the cause and its effects are based on estimates of unconditional contingencies. Deterministic common-cause models imply that all effects should be present if there is unambiguous evidence for the presence of the cause. This inference is invited even when there is no direct evidence for the presence of some of the effects. Thus, in diagnostic blocking tasks, participants are expected to infer that both effects of the common cause were already produced in earlier learning phases even though the second effect was only explicitly mentioned in the later phase. This implication of causal-model theory was tested and supported in Experiment 2.

Another important feature of causal-model theory is its claim that participants are able to access their causal representations either in the predictive cause-effect or the diagnostic effect-cause direction, depending on the question posed. This claim has often been overlooked (Matute et al., 1996) or misunderstood (Shanks & Lopez, 1996; see also Waldmann & Holyoak, 1997). Part of the confusion may be due to the fact that the empirical evidence for this claim in Waldmann and Holyoak (1992) was based on a cross-experiment comparison. Experiment 3a presents a direct test of this claim. A condition in which an effect cue was potentially caused by one cause was compared with an effect cue that could be caused by two different causes. Participants clearly proved sensitive to this difference, giving lower ratings to the ambiguous effect.

The predictions of causal-model theory regarding predictive learning were also tested in a more detailed fashion. Using a novel blocking paradigm, Experiment 3a compared a blocking situation in which a redundant cause is blocked by two already established causes with a task in which it is blocked by only one cause. Associative theories predict complete blocking provided that Phase 1 learning proceeded to the asymptote. According to causal-model theory blocking is based on the assessment of conditional contingencies. Blocking is expected to be partial when the relevant conditional contingencies are unavailable and when the cause-effect relation in Phase 1 is deterministic. Furthermore, a difference between the two conditions is predicted, as the ambiguous cue condition provides more relevant information about conditional contingencies. Experiment 3a supports causal-model theory. Experiment 3b tests this theory against the alternative associationist theory that partial blocking and the difference between the two conditions are due to preasymptotic learning in Phase 1. Contradicting this account, no evidence for pre-asymptotic learning was observed. Interestingly, two presentations of each case were sufficient to generate asymptotic rating patterns. This result is more in line with theories that assume contingency learning than with theories that postulate incremental changes of associative weights.

The lack of a learning curve in Experiment 3b is at odds with previous demonstrations of learning curves with probabilistic structures (Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Chatlosh, Neunaber, & Wasserman, 1985; Shanks, 1987) although not all studies managed to provide evidence for their existence (Baker, Berbrier, & Vallée-Tourangeau, 1989, Experiment 3). According to associationist accounts a few trials indicating a deterministic relation should be interpreted in the same way as many trials indicating a probabilistic relation as long as the corresponding associative weights are equal. This counterintuitive prediction was clearly refuted by Experiment 3b. Participants needed no more than two presentations of each case to be fairly sure about the relation between the causes and effects. Even though causal-model theory does not claim to be a process theory of causal learning, a possible reconciliation between the findings in the deterministic and the probabilistic case may be suggested. Deterministic hypotheses can be refuted with single cases whereas the assessment of a probabilistic causal relation clearly requires more trials to become stable. Thus, participants may be more conservative in their causal assessments when they observe probabilistic relations in the early stages of learning than later (see also Baker et al., 1996, for a similar explanation of learning curves).

Discussion of Critical Results of Proponents of Associative Learning Theories

This article presents four experiments that clearly favor causal-model theory. This raises the question of how causal-model theory would handle the empirical results that have been proffered to refute it. It is important to note that causal-model theory only makes predictions about causal learning. It is not meant to be a model of arbitrary associative learning (see also Waldmann & Holyoak, 1997). This caveat has to be kept in mind when we evaluate apparent evidence against causal-model theory. A number of studies have shown cue competition with cues that may be interpreted as effects (e.g., Shanks, 1991; Shanks & Lopez,

1996). However, simply presenting symptoms of diseases as cues does not guarantee that participants interpret these cues as effects of a common cause, a prerequisite for the absence of diagnostic blocking. As pointed out by Waldmann and Holyoak (1992), symptoms may be causes of a disease or play other causal roles within a complex causal network (see also Patel & Groen, 1986).

A different possibility is raised by Price and Yates's (1995, Experiment 4) recent failure to find any evidence for asymmetries of blocking in a predictive and a diagnostic condition using a one-stage blocking design. In their experiment, participants were presented with light switches (causes) or indicator lights (effects) that were probabilistically related to the state of a nuclear power plant. Apart from the possibility that participants may find it implausible that a nuclear power plant has probabilistic switches or indicator lights, the probabilistic nature of this task may be a key factor in explaining the difference from the present findings (and from Waldmann & Holyoak, 1992, Experiment 3). It is obvious that learning about deterministic relations is easier than learning about probabilistic relations. Furthermore, diagnostic learning seems to be more difficult than predictive learning (see Bindra, Clarke, & Shultz, 1980). Diagnostic learning involves retrospective revisions of an already formed causal model on the basis of new effect information. In contrast, predictive learning permits the successive updating of a mental model parallel to the unfolding of the events in the world. A further complication in many studies of diagnostic learning is that they often require the learner to make assumptions about the existence of causal events that are not explicitly mentioned (see the present Experiment 2 for an example). Thus, it is plausible to expect that the competence of humans to learn about diagnostic causal relations and to give adequate diagnostic assessments may sometimes break down when the task is too complex. It may well be that in Price and Yates's (1995) study participants imposed a predictive interpretation in both situations according to which cues were seen as causal predictors of the state of the power plant.

In general, causal-model theory is presently restricted to model people's competence to learn about causal structures. This competence displays itself best when potential information processing constraints are reduced. The present experiments show that people have the capacity to ignore the order in which the learning input is presented and are capable of transforming learning events into adequate representations of causal models that honor causal directionality. However, it seems likely that with more complex tasks this competence may partly break down, and additional factors such as learning order may start to play a role.

Another problem with translating designs from noncausal, associative learning paradigms into causal learning tasks is the possibility of creating ecologically implausible causal situations. For example, Shanks and Lopez (1996, Experiment 3) presented participants with the following trial types: (1) Cause → Effect A and Effect B, (2) Cause → Effect C, (3) No-Cause → Effect B. This rather unusual structure describes a causal situation with disjunctive effects. The cause (a fictitious disease) either is followed by the effect complex A and B or by the effect C but no other combination of these three potential effects is observed. This learning input is clearly inconsistent with a common-cause model in which a common cause with stable properties independently generates three effects (see Waldmann & Holyoak, 1997, for a more detailed critique of Shanks & Lopez, 1996). Miller and Matute (1998) present a different example of this problem. In one condition of their experiment, a cause produced a specific effect O1 on the first five days, whereas on the sixth training day a second effect O2 suddenly appeared along with O1. Again, this is a situation which is incompatible with a common-cause structure in which the causal power of the cause remains stable across learning phases. It seems likely that learners confronted with an implausible, complex causal structure will tend to ignore the causal character of the instructions.

Matute et al. (1996) criticized causal-model theory on the basis of apparent evidence for competition among effects from their experiments employing one-phase blocking tasks in which cues either were causes or effects. When they asked whether the critical (potentially blocked) effect cue was an effect of the cause or whether the cause produced this effect, then they found (perfectly in line with causal-model theory) no evidence for effect competition. When in their Experiment 3, however, the test question asked whether the cue is "indicative" of the cause, then participants tended to express the validity of the cue relative to the strength of collateral effects in their ratings.

From the perspective of causal-model theory the results of this study show that some test questions elicit judgments that express the validity of a cue relative to the validity of other cues. A single symptom of a disease may be rated as more indicative if it stands out as opposed to being one of many equally valid signs. However, the fact that we sometimes give relative judgments does not necessarily imply that our learning about the absolute strengths of different cause-effect links underlying the relative judgments is also affected. According to causal-model theory it is necessary to distinguish between competition at retrieval that is based on relative judgments of established different causal links and competition at learning that is based on the impossibility to access the relevant conditional contingencies to decide upon the status of a potential cause.

Associative Learning Versus Updating of Causal Models

Sensitivity to causal directionality points to a more fundamental difference hidden beneath the controversy about associative theories and causal-model theory. Most associative theories still embody the neobehaviorist view of stimulus-bound learners being confronted with cues and predicting outcomes that unfold in parallel with the learning process. Cognitive representations are just intermediate steps in the association between stimuli and responses. This account works fairly well as long as the learning task conforms to this paradigmatic situation, but fails when more complex tasks are studied.

Diagnostic learning is a task in which the sequence of the

observed events is decoupled from the sequence of events in the real world. The observation of effect cues should lead to diagnostic inferences about events (causes) that occurred earlier in time but were not actively encoded. Thus, diagnostic learning is a test case for humans' competency to form and update mental representation in the absence of direct stimulation. Causal-model theory is an account that attempts to separate mental representations from the arbitrary sequence of learning events. Cues need not necessarily be assigned to the input layer of mental models. They are rather assigned on the basis of their causal role within the causal network that represents physical reality. The directionality embodied within mental representations may be completely independent from the directionality of learning events. How learning events are interpreted is a joint function of prior knowledge and the structure of the learning input (see also Waldmann, 1996). As yet very little is known about the details of the mechanisms involved in the updating of causal representations (but see Waldmann & Martignon, 1998) and the processing constraints that place limits on people's competence.

References

Baker, A. G., Berbrier, M., & Vallée-Tourangeau, F. (1989). Judgments of a 2 × 2 contingency table: Sequential processing and the learning curve. *Quarterly Journal of Experimental Psychology, 41B,* 65–97.

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

Baker, A. G., Mercier, P., Vallée-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 414–432.

Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 1–45). San Diego, CA: Academic Press.

Bindra, D., Clarke, K. A., & Shultz, T. R. (1980). Understanding predictive relations of necessity and sufficiency in formally equivalent "causal" and "logical" problems. *Journal of Experimental Psychology: General, 109,* 422–443.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition, 18,* 537–545.

Chatlosh, D. L., Neunaber, D. J., & Wasserman, E. A. (1985). Response-outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation, 16,* 1–34.

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

Connolly, T. (1977). Cues, components and causal structure in laboratory judgment studies. *Educational and Psychological Measurement, 37,* 877–888.

Connolly, T., & Srivastava, J. (1995). Cues and components in multiattribute evaluation. *Organizational Behavior and Human Decision Processes, 64,* 219–228.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 49*(B), 60–80.

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

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

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 276–296). New York: Appleton-Century-Crofts.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Matute, H., Arcediano, F., & Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 182–196.

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, and Cognition, 19,* 1398–1410.

Miller, R. R., & Matute, H. (1998). Competition between outcomes. *Psychological Science, 9,* 146–149.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation. Advances in research and theory* (Vol. 22, pp. 51–92). New York: Academic Press.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10,* 91–116.

Pearce, J. M., & Hall, G. (1980). A model of Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

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

Price, P. C., & Yates, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1639–1655.

Reichenbach, H. (1956). *The direction of time.* Berkeley, CA: 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. (1995). Full preservation of a response-outcome association through training with a second outcome. *The Quarterly Journal of Experimental Psychology, 48B,* 252–261.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Shanks, D. R. (1987). Acquisition functions in contingency judgment. *Learning and Motivation, 18,* 147–166.

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

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 229–261). New York: Academic Press.

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

Tassoni, C. J. (1995). The least mean squares network with information coding: A model of cue learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 193–204.

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 intervent relations: From cause to effect and from effect to cause. *Memory & Cognition, 21,* 802–808.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25,* 127–151.

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: Vol. 34. Causal learning* (pp. 47–88). San Diego, CA: Academic Press.

Waldmann, M. R., & Holyoak, K. J. (1990). Can causal induction be reduced to associative learning? *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. (1997). Determining whether causal order affects cue selection in human contingency learning: Comments on Shanks and Lopez (1996). *Memory & Cognition, 25,* 125–134.

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.

Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1102–1107). Mahwah, NJ: Erlbaum.

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, 4,* 96–194.

Williams, D. A., Sagness, K. E., & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 694–709.

Appendix

Derivation of Asymptotic Weights for the Ambiguous and Unambiguous Cue Conditions

In this appendix the asymptotic weights for the redundant cues in the unambiguous and ambiguous cue conditions (Experiment 3) are derived. Of particular interest are situations in which the predictive cues were not trained to their asymptote within Phase 1. The following derivations are based on the Rescorla–Wagner rule,

$$\Delta V_{fc} = \alpha_f \cdot \beta_c \cdot \left(\lambda - \sum_f V_{fc}\right),$$

where ΔV_{fc} represents the weight change of a particular cue f with respect to the output category c. The learning parameter α_f is a measure of the saliency of this cue, and β_c represents the saliency of the output category that occurs on the respective trial. λ is the asymptote of associative strength (typically 1), and

$$\sum_f V_{fc}$$

is the sum of the associative weights of all the cues and the category that are present on a particular trial. When the category c is absent on a particular trial, λ is set to the value 0 and the sum term represents the sum of the associative weights of the cues that are present in the absence of the category. In this situation a reduction of the weights occurs (i.e., extinction).

The Rescorla–Wagner rule is equivalent to the least-mean-squares rule (LMS rule) of Widrow and Hoff (1960). This rule implements an iterative algorithm to solve a set of linear equations defined by the set of stimulus-response patterns presented in the learning phases. The following derivations are based on the LMS rule in which, for convenience, the different learning rates are assumed to be equal to a fixed μ.

The LMS learning rule minimizes the sum of squared error over the presented stimulus patterns. That is, the error function

$$E(V(t)) = \frac{1}{2}\sum_{p \in P}\left(\lambda_p - \sum_f V_{fcp}(t)\right)^2 \qquad \text{(A1)}$$

will be minimized. In this equation $E(V(t))$ represents the error that a specific set of weights V generates at a specific point in time t, and p stands for a specific training pattern from the set of all patterns P. $\Sigma V_{fcp}(t)$ is an expression for the sum of all associative weights of cues that are present in a particular pattern with respect to the output category at a specific point in time.

Phase 1: Asymptotic Weights

Phase 1 is identical across all conditions. Three types of cues are presented. The predictive cue, *A1*, is a perfect predictor of one of the outcomes (e.g., Midosis), *B1* is uncorrelated with either outcome, and *C1* is perfectly correlated with the second outcome. Because in the unambiguous cue condition only the outcome that is paired with the redundant cue is of interest, and because the ambiguous cue condition is symmetric, the following derivations focus on only one outcome.

For the patterns within Phase 1 the following error function can be derived:

$$E(V(t)) = \frac{1}{2}[(1 - V_{A1}(t))^2 + (0 - V_{B1}(t))^2 + (0 - V_{C1}(t))^2].$$

This equation is an instantiation of the general error function (Equation 1). The first squared component expresses the error that is made when cue *A1* is present and is supposed to predict the presence of the corresponding outcome ($\lambda = 1$). When the cues *B1* and *C1* are present the desired prediction is the absence of this outcome. Hence the λ is set to 0 for these cues. The weight changes that minimize the error function can be obtained by computing the partial derivatives with respect to each weight:

$$\Delta V_{A1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{A1}(t)} = \mu(1 - V_{A1}(t))$$

$$\Delta V_{B1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{B1}(t)} = -\mu V_{B1}(t)$$

$$\Delta V_{C1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{C1}(t)} = -\mu V_{C1}(t).$$

The asymptotic weights can be obtained by setting these partial derivatives to 0. At the end of the learning process, when t approaches t_e, the weights are

$$V_{A1}(t_e) = 1 \text{ and } V_{B1}(t_e) = V_{C1}(t_e) = 0.$$

V_{A1} monotonically approaches 1, whereas V_{B1} and V_{C1} stay at 0.

Phase 2: Unambiguous Cue Condition

The following error function can be derived for the learning patterns of the unambiguous cue condition in Phase 2. In the unambiguous cue condition an additional redundant cue, *D1*, is redundantly paired with *A1* to predict the presence of the outcome.

$$E(V(t)) = \frac{1}{2}[(1 - V_{A1}(t) - V_{D1}(t))^2 + (0 - V_{B1}(t))^2 + (0 - V_{C1}(t))^2].$$

The weight changes are expressed by

$$\Delta V_{A1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{A1}(t)} = -\mu(1 - V_{A1}(t) - V_{D1}(t))$$

$$\Delta V_{B1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{B1}(t)} = -\mu V_{B1}(t)$$

$$\Delta V_{C1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{C1}(t)} = -\mu V_{C1}(t)$$

$$\Delta V_{D1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{D1}(t)} = -\mu(1 - V_{A1}(t) - V_{D1}(t)).$$

It can be seen that at each point in time t the weight changes for the predictive and the redundant cues are equal: $\Delta V_{A1}(t) = \Delta V_{D1}(t)$. Let t_0 represent the end of Phase 1. $V_{A1}(t_0)$ represents the associative weight of the predictive cue *A1* at the end of Phase 1. Thus, $V_{A1}(t_0) + \Delta V_{A1} + \Delta V_{D1} = 1$, where ΔV_{A1} and ΔV_{D1} represent the maximal weight gains of these two cues within Phase 2. Generally the weight change occurring between t_0 and t_e of the predictive cue *A1* and the redundant cue can be expressed by

$$\Delta V_{A1} = \Delta V_{D1} = \frac{1}{2}(1 - V_{A1}(t_0)).$$

Hence blocking is complete for cue *D1* when *A1* is trained to the asymptote in Phase 1. When *A1* is preasymptotic at the end of Phase 1, then the difference to the asymptote is equally divided between the two cues in Phase 2.

Phase 2: Ambiguous Cue Condition

In this condition the redundant cue *D1* is paired with *A1* to predict one of the outcomes (e.g., Midosis), but, unlike the unambiguous condition, it is also paired with the second predictive cue *C1* to predict the alternative outcome (e.g., Xeritis). Thus, in this trial *C1* and *D1* are present in the absence of the first outcome (Midosis). Therefore the following slightly different error function applies to the Phase 2 patterns of the ambiguous cue condition:

$$E(V(t)) = \frac{1}{2}[(1 - V_{A1}(t) - V_{D1}(t))^2 + (0 - V_{B1}(t))^2 + (0 - V_{C1}(t) - V_{D1}(t))^2].$$

For the different cues the following weight changes can be derived:

$$\Delta V_{A1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{A1}(t)} = -\mu(1 - V_{A1}(t) - V_{D1}(t))$$

$$\Delta V_{B1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{B1}(t)} = -\mu V_{B1}(t)$$

$$\Delta V_{C1}(t) = -\mu \frac{\partial E(V(t))}{\partial V_{C1}(t)} = -\mu(V_{C1}(t) + V_{D1}(t))$$

$$\Delta V_{D1}(t) = -\mu \left(\frac{\partial E(V(t))}{\partial V_{A1}(t)} + \frac{\partial E(V(t))}{\partial V_{C1}(t)} \right).$$

The weights at t_e when the error function is at its minimum can again be computed by setting the partial derivatives with respect to each weight to 0. This operation yields the following relations between the cues:

$$V_{A1}(t_e) + V_{D1}(t_e) = 1, \quad V_{C1}(t_e) + V_{D1}(t_e) = 0, \quad V_{B1}(t_e) = 0. \quad \text{(A2)}$$

At the end of Phase 1, that is at t_0, the weights of the cues are

$$V_{A1}(t_0) = a, \quad V_{B1}(t_0) = V_{C1}(t_0) = V_{D1}(t_0) = 0.$$

These are the values of the associative weights at the beginning of Phase 2 (a may be preasymptotic). Then the vector $\vec{V}_{*1}(t_e)$ represents the weights at the end of Phase 2 training:

$$\vec{V}_{*1}(t_e) = \begin{pmatrix} a + \Delta V_{A1} \\ \Delta V_{B1} \\ \Delta V_{C1} \\ \Delta V_{D1} \end{pmatrix} = \begin{pmatrix} a + \Delta V_{A1} \\ 0 \\ a + \Delta V_{A1} - 1 \\ 1 - a - \Delta V_{A1} \end{pmatrix}.$$

Because ΔV_{A1} is variable, the vector describes the one-dimensional solution space for Equation A2. Because of the gradient descent method implemented by the LMS rule the solution will be computed that is closest to the starting vector. This solution can be found by setting the derivative of the absolute value of the difference vector $\Delta\vec{V}_{*1}$ of the weights from the beginning and the end of Phase 2 to 0.

$$\|\Delta\vec{V}_{*1}\|_2 = \|\vec{V}_{*1}(t_e) - \vec{V}_{*1}(t_0)\|_2 = \left\| \begin{pmatrix} \Delta V_{A1} \\ 0 \\ a + \Delta V_{A1} - 1 \\ 1 - a - \Delta V_{A1} \end{pmatrix} \right\|_2$$

$$\|\Delta\vec{V}_{*1}\|_2 = \sqrt{\Delta V_{A1}^2 + (a + \Delta V_{A1} - 1)^2 + (1 - a - \Delta V_{A1})^2}$$

$$\frac{d\,\|\Delta\vec{V}_{*1}\|_2}{d\,\Delta V_{A1}} = \frac{1}{2} \cdot \frac{2\Delta V_{A1} + 4(a + \Delta V_{A1} - 1)}{\sqrt{\Delta V_{A1}^2 + 2(a + \Delta V_{A1} - 1)^2}} \stackrel{!}{=} 0$$

Eventually we obtain

$$\Delta V_{A1} = \frac{2}{3}(1 - a), \quad \Delta V_{D1} = \frac{1}{3}(1 - a).$$

Or,

$$\Delta V_{A1} = \frac{2}{3}(1 - V_{A1}(t_0)), \quad \Delta V_{D1} = \frac{1}{3}(1 - V_{A1}(t_0)).$$

Therefore, in situations in which the predictive cue is not trained to its asymptote within Phase 1, the redundant cue grows more slowly and the predictive cue grows faster in the ambiguous cue condition than in the unambiguous cue condition in Phase 2. Whereas in the ambiguous cue condition the predictive cue rises twice as fast as the redundant cue, in the unambiguous condition both cues approach the asymptote at equal pace.

Received October 4, 1996
Revision received June 3, 1999
Accepted June 14, 1999 ■