Learning & Behavior
2005, 33 (2), 211-229

Competence and performance in causal learning

MICHAEL R. WALDMANN
University of Göttingen, Göttingen, Germany

and

JESSICA M. WALKER
University of California, Los Angeles, California

The dominant theoretical approach to causal learning postulates the acquisition of associative weights between cues and outcomes. This reduction of causal induction to associative learning implies that learners are insensitive to important characteristics of causality, such as the inherent directionality between causes and effects. An ongoing debate centers on the question of whether causal learning is sensitive to causal directionality (as is postulated by causal-model theory) or whether it neglects this important feature of the physical world (as implied by associationist theories). Three experiments using different cue competition paradigms are reported that demonstrate the competence of human learners to differentiate between predictive and diagnostic learning. However, the experiments also show that this competence displays itself best in learning situations with few processing demands and with convincingly conveyed causal structures. The study provides evidence for the necessity to distinguish between competence and performance in causal learning.

The dominant theoretical approach to causal learning postulates the acquisition of associative weights between cues and outcomes. These associative weights reflect the amount of covariation between the learning events. In the past few years, the associationist approach to causal learning has been criticized by a number of researchers (e.g., Cheng, 1997; De Houwer & Beckers, 2002; Glymour, 2001, 2003; Gopnik et al., 2004; Lovibond, 2003; Waldmann, 1996, 2000, 2001; Waldmann & Hagmayer, 2005; Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne, 1995). The main thrust of this critique is that causality cannot be reduced to mere covariation detection. Cues and outcomes may covary either because they are *causally* related, or because they are *spuriously* related, an important distinction to which associationist models are not sensitive. In addition, there are cases in which no statistical covariation can be measured (e.g., when the base rate of the effect is already at the ceiling in the absence of the cause), but it is nevertheless not appropriate to infer that the cause is not effective (see Wu & Cheng, 1999). Finally, covariations are inherently symmetric and therefore insensitive to the crucial property of *causal directionality*: Causes influence their effects but not vice versa. Knowledge about the direction of causality guides our actions. Causes can be set to achieve effects whereas effects do not generate their causes. Moreover, patterns of causal directionality have statistical implications that a learning mechanism could capitalize on (see Waldmann et al., 1995). For example, multiple effects of a common cause (common-cause model) are (spuriously) correlated but become independent conditional upon the states of their cause, whereas multiple causes of a common effect (common-effect model) tend to be independent (unless they are part of a larger network) but become dependent when the common effect is held constant (see Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). These structural differences have consequences for how causal power normatively should be assessed, but are neglected by associative theories that typically impose the same type of network on learning events irrespective of their causal roles (Waldmann, 2000; Waldmann & Hagmayer, 2001).

Sensitivity to Causal Directionality: Current Controversies

In the past few years, a number of researchers have addressed the question of whether human learners are sensitive to the aspect of causal directionality in learning. Waldmann and Holyoak (1992) have proposed a *causal-model theory*, which postulates that prior knowledge about causal directionality guides processing of the learning input in causal induction. In a number of experiments, they have observed that human learners are indeed sensitive to causal directionality, and are able to differentiate between the order in which information is

Experiments 1 and 3 were conducted while M.R.W. was affiliated with the Max Planck Institute for Psychological Research, Munich. Experiment 2 was planned in collaboration with J.M.W. during a sabbatical of M.R.W. from the University of Göttingen at the Department of Psychology of the University of California, Los Angeles. The research was supported by a DFG grant (Wa 621/5-2). We thank L. Allan, P. Cheng, J. De Houwer, K. Holyoak, and J. Tangen for helpful comments. Portions of this research were presented by M.R.W. at the invited symposium "Is Everyday Causal Reasoning Rational?" of the 2000 Annual Meeting of the Psychonomic Society, New Orleans. Correspondence should be sent to M. Waldmann, Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany (e-mail: michael.waldmann@bio.uni-goettingen.de).

 Copyright 2005 Psychonomic Society, Inc.

presented within learning trials and the order of events in the real world to which the trials are referring. However, the conclusions drawn from these initial studies have been hotly debated, and the conditions under which causal learning is directional remain a focus of controversy.

The debate was initiated by Waldmann and Holyoak's (1992) experiments in which a two-phase blocking paradigm was used. Participants learned in Phase 1 that a predictive cue, for example a light button (Experiment 3), is perfectly correlated with the outcome, in this example the state of the alarm in a bank. In Phase 2, an additional button, previously not mentioned, was redundantly paired with the predictive cue from Phase 1. Now, whenever both buttons were on, the alarm was on, and when they both were off, the alarm also was off. In the test phase, participants were asked to rate how predictive of the state of the alarm each button was individually. The crucial manipulation involved the initial cover stories. In a *predictive-learning condition*, the buttons were described as potential causes of the alarm (the effect), whereas in the *diagnostic-learning condition* the buttons were characterized as effects of the alarm (the cause). Thus, in the predictive-learning condition, participants learned about a common-effect model, whereas in the diagnostic-learning condition, identical learning trials were used to acquire a common-cause model.

Only the cover stories varied in this design; the learning trials and test questions were identical across both conditions. Accordingly, associative theories such as the Rescorla–Wagner theory (Rescorla & Wagner, 1972) predict blocking of the redundant cue in both conditions. According to this learning rule, learning takes place only when something unexpected happens. Since the predictive cue from Phase 1 perfectly predicts the outcome in both phases, there is no reason to learn anything about the redundant cue in Phase 2.

Contrary to this prediction, Waldmann and Holyoak (1992) found that the blocking effect interacted with the causal status of cues and outcomes: Blocking was observed only in the predictive but not in the diagnostic condition. This effect is predicted by causal-model theory (Waldmann, 1996, 2000, 2001; Waldmann & Holyoak, 1992), which postulates that assumptions about abstract causal models interact with the processing of the learning input. In the predictive condition, the cues are assigned the role of potential causes, and the outcome the role of a common effect. Assessing causal strength within common-effect models requires holding constant the potential influence of alternative causes. A typical feature of the blocking paradigm is that the redundant cue can never be observed in the *absence* of the predictive cue, which makes it impossible to assess the individual causal power of the redundant cue. Although the redundant cue can be observed in the *presence* of the alternative cue, this cue represents a deterministic cause that creates a ceiling effect so that the potential additional impact of the redundant cue cannot possibly display itself (see also Cheng, 1997; Wu & Cheng, 1999). Both factors should lead to assessments that express uncertainty about the causal status of the redundant cue. Thus, unlike associative theories, which predict full blocking of the redundant cue (i.e., certainty that it is not a cause) in case Phase 1 learning proceeded to the asymptote, causal-model theory predicts partial blocking (i.e., uncertainty rather than certainty about the redundant cue). Previous experiments support causal-model theory. Participants typically express uncertainty in their ratings (Waldmann, 2000, 2001; Waldmann & Holyoak, 1992). Only with effects with probabilities below the ceiling should full blocking be observed. This prediction is supported by recent findings with continuous effect variables that showed that the blocking effect in predictive learning is stronger when the predictive cue causes an effect at a submaximal intensity as compared with a maximal-intensity condition (De Houwer, Beckers, & Glautier, 2002).[1] In the submaximal scenario, the effect is not at the ceiling in Phase 1, which allows the additional redundant cue to display its potential causal impact.

By contrast, in the diagnostic condition, the cues are assigned the role of potential effects of a common cause. Assessing causal strength within a common-cause model does not require holding constant alternative effects. Thus, participants should have learned that the common cause has two deterministic effects. Since no alternative causes of these effects were mentioned, both effects should be rated as equally diagnostic for their common cause (i.e., blocking would be absent).

Complete absence of a descriptive blocking effect (i.e., equal ratings for predictive and redundant cues) is, even within the framework of causal-model theory, only predicted in specific situations. Waldmann (2000, Experiment 2) has shown that in the diagnostic-learning condition, participants need to believe that the causal model did not change between the learning phases. When they learn about the redundant effect in Phase 2, they should infer that this effect had already been produced by the common cause in Phase 1, although no information was given about its presence. If some participants believed that the redundant cue was absent in Phase 1, different ratings would be predicted. Similarly, De Houwer (2002) has shown for a predictive-learning task that the size of the blocking effect is dependent on whether participants believed that the redundant cause was absent in Phase 1 or whether they retrospectively inferred its presence. Only in the first condition was a blocking effect seen. These effects are normative because the contingencies underlying the judgments vary with different assumptions about presence and absence of cues in previous learning phases.

A second factor that normatively affects the blocking effect is domain knowledge that might affect judgments. Waldmann (2001) has shown in the diagnostic condition of an overshadowing paradigm that the redundant cues were rated slightly lower than the predictive cue when the task was about diseases. The difference between predictive and diagnostic learning was still highly significant but the pattern was slightly different from the results of a second study that used the same learning trials but instead presented an artificial device as a learning domain (see also Waldmann, 2000, for similar descriptive patterns in blocking experiments). Waldmann (2001) argued that in the real world, diseases unlike devices, typically are embedded in complex causal networks with hidden causes (open-world scenarios). The flu causes fever but fever can occur because of any number of alternative, in part unknown, causes. In these domains, causal-model theory and other normative theories predict lowered ratings for the redundant cue. In blocking paradigms, the redundant symptom is always shown together with the predictive symptom as a sign of the disease. If in the test phase participants rate the redundant symptom by itself, they seem to make the assumption that this symptom individually might be caused by other, unknown, causes, and therefore conservatively lower the ratings slightly. This does not happen with devices because hidden causes are unlikely. Thus, on the assumption that people are affected by abstract domain knowledge, the observation of a small blocking effect in diagnostic-learning conditions is consistent with causal-model theory.

Waldmann and Holyoak's (1992) initial experiments have faced a number of critiques. Some critics questioned the reliability of the effect (Cobos, López, Cano, Alvarez, & Shanks, 2002; Matute, Arcediano, & Miller, 1996, Experiment 3; Price & Yates, 1995; Shanks & López, 1996), but in the meantime the basic finding has been replicated in several additional experiments that used different cue competition paradigms (Matute et al., 1996, Experiments 1, 2; Tangen & Allan, 2004; Tangen, Allan, & Sadeghi, 2005; Van Hamme, Kao, & Wasserman, 1993; Waldmann, 2000, 2001). Tangen et al.'s study is particularly interesting because it extends the findings supporting causal-model theory to domains with probabilistic relations.

Boundary Conditions for Causal Learning

Every theoretical account has to presuppose some boundary conditions for the postulated learning processes. For example, it is unlikely that associative learning mechanisms are unaffected by performance-limiting factors such as the number of cues and outcomes and the complexity of the causal relations. These restrictions are rarely stated in the exposition of the theory but are part of implicit knowledge guiding the design of experiments. It seems plausible that human causal learning operates optimally only in restricted circumstances. In principle, causal models could be postulated for arbitrarily complex models, but it seems obvious that very soon information processing unaided by mechanical tools would face capacity restrictions that limit the competence to correctly learn and reason with these models (see also Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). As a matter of fact, partly because of these limitations, a number of statistical tools have recently been developed to aid researchers in the analysis of complex causal structures (e.g., structural equation modeling).

Causal-model theory can be used to analyze the potential difficulties learners face with complex causal learning tasks. The predictive as well as the diagnostic component might be affected. Predictive learning from multiple causes to a common effect requires holding cofactors constant when learners are assessing the strength of a target relation. Unlike most associative theories, which have cue competition built into the learning mechanism (e.g., Rescorla & Wagner, 1972), causal-model theory postulates a higher level process of selecting relevant cofactors and of computing statistical contingencies within subsets of the event space in which the states of the cofactors are held constant (see Hagmayer & Waldmann, 2002; Waldmann, 1996; Waldmann & Hagmayer, 2001; Waldmann & Martignon, 1998). For generative causes, the cases are most informative in which alternative causes are absent. Estimating conditional contingencies is not too demanding when the target cause frequently occurs in the absence of alternative causes because learners can then easily focus on the most informative segments of the data. If the presence and absence of the cofactors are frequent and intermixed in the learning trials, it should be harder to separately store the informative cases in memory. Waldmann and Hagmayer (2001) have indeed shown that participants tended to control for the cofactor only in a predictive trial-by-trial learning task with a complex statistical structure (Simpson's paradox) when the trials were blocked according to different states of the cofactor. When the cases were intermixed, however, the cofactor was neglected by many learners.

In the blocking paradigm, participants face an extreme case of a potential cause whose observed relation with the effect is perfectly confounded by an alternative cause within every single learning trial. Therefore, learners are confronted with two conflicting cues. Unconditionally, the redundant cue is perfectly correlated with the outcome. However, information about the conditional contingency in the absence of the confound is not available, which normatively should lead to the conclusion that the information is insufficient to warrant any conclusions. The redundant cue may be an individual cause, it may interact with the predictive cue, or it may just be a spurious correlate of the cause. Because of the conflict between the cues, it is predicted that, under circumstances of reduced capacity, learners should have difficulties with grasping this situation, and rather fall back on the simpler, salient cue of the perfect contingency within Phase 2. Thus, a *reduction* of the blocking effect is predicted with increasing demands on information processing capacity.

This prediction is supported by a study by De Houwer and Beckers (2003). They showed that the blocking effect decreased when participants were presented with a difficult secondary tone-counting task as compared with an easy secondary task. The difference was observed only when the secondary task was presented in both the learning and test phases, not when the test phase was spared. However, their interpretation of their finding of reduced blocking with increasing load differed from ours. They argued that full blocking relies on the inference that the redundant cue did not add anything to the effect already caused by the predictive cause. Increased processing load might interfere with this inference, which should lead to uncertainty (i.e., partial blocking in our terminology) as opposed to full blocking (see note 1).

Diagnostic learning should also be affected by capacity limitations, but for different reasons. Common-cause models do not require holding alternative effects constant. However, diagnostic learning requires other specific capacity-demanding processes that differentiate it from predictive learning. Originally, causal-model theory was developed to model the *competence* of causal learning. Specifically, it was assumed that people learn about cause–effect contingencies independently of the order of learning events. For example, when a common-cause model is being acquired, it should (normatively and according to causal-model theory) not matter whether learning proceeds in the cause–effect or the effect–cause direction. However, psychologically it seems implausible that learning order does not play a role unless the learning situations are fairly simple. In diagnostic tasks, effect information is presented first (e.g., the symptom of a disease); then later, after the diagnostic judgments, learners receive feedback about the cause (the disease) that produced the earlier observed effects. Thus, the sequence of learning events runs opposite to the sequence of the real events that underlie them. We know that in the real world, causes precede their effects, irrespective of the order in which they are presented to us. Therefore, the learning events need to be mentally reorganized to lead to a causal model that is comparable with causal models acquired in the predictive direction (from causes to effects).

Preliminary evidence for the assumption that mismatches between presentation order and real-world temporal order tax working memory comes from a text comprehension study by Münte, Schiltz, and Kutas (1998). These investigators showed that reading sentences in which the order in which events were mentioned in the text runs opposite to the order of these events in the described real-world situation (e.g., "Before the scientist submitted the paper, the journal changed its policy") required more working memory than reading sentences in which the two orders corresponded (e.g., "After the scientist submitted the paper, the journal changed its policy"). The neural correlate of working memory in the prefrontal cortex was measured with electrophysiological methods (event-related potentials). These results encourage the hypothesis that diagnostic learning will impose higher working-memory demands than predictive learning.

The hypothesis of greater demands of diagnostic reasoning is also supported by recent experiments by Fenker, Waldmann, and Holyoak (in press), who investigated a semantic memory task. In the experiments, participants received word pairs consecutively (e.g., spark–fire) and had to assess as quickly as possible whether the word pair referred to causally related events or not. With word pairs that were equally associated in both directions, the results showed that it takes longer to check whether a diagnostic effect–cause relation (fire–spark) is true than whether a predictive cause–effect relation (spark–fire) is true. No difference in reaction times was observed when participants had to assess whether the words were merely associated. These results provide evidence for the special status of causal relations and for the greater difficulty of accessing diagnostic knowledge.

A further asymmetry between predictive and diagnostic learning is the relatively greater difficulty of updating causal strength estimates in diagnostic learning. According to causal-model theory, learning involves the updating of conditional probabilities used to estimate causal strength. In predictive learning, the cause information is given first, which allows the learner to select the relevant conditional probability estimates that are going to be updated. In diagnostic learning, the effect information is presented first. Again, conditional probabilities that are directed from causes to effects are going to be updated (e.g., $P(\text{effect}_1|\text{cause}_1)$, see Waldmann & Martignon, 1998). However, in this situation the information about the particular effect has to be stored until after feedback about the corresponding cause is given. An observed effect may be produced by alternative causes so that it is not clear at first which estimate is going to be updated. This is a consequence of the inherent asymmetry between causes and effects. Multiple causes of an observed effect compete, whereas multiple effects of a common cause do not. Thus, the greater difficulty of assessing causal strength in diagnostic-learning situations provides a further reason for expecting that normatively appropriate diagnostic learning may be found only in situations with relatively little complexity.

A final reason for difficulties with diagnostic learning may be derived from evolutionary considerations. In our natural environment, we typically perceive events in their natural causal order. Causes precede effects, and therefore our learning events are typically also ordered in this direction. There is some evidence that animals are able to reason backward in time (Esmoris-Arranz, Miller, & Matute, 1997), but it is well known that this competence is harder to demonstrate than forward reasoning. Also, there is evidence from studies with children showing that backward memory–driven inferences are harder to accomplish than reasoning processes that proceed in line with the learning direction (Bindra, Clarke, & Shultz, 1980). These studies typically presented causes before effects in the learning phase, but in diagnostic-learning tasks the problems with reasoning from effects to causes are exacerbated because participants perceive cause information after effect information, which experientially contradicts the natural order of events.

It is important to note that diagnostic reasoning and diagnostic learning are two different notions. It is possible to learn a cause–effect relation in the predictive order and then reason diagnostically back in time. This is different from diagnostic learning in which effect cues are experienced prior to cause outcomes, although the events in the real world to which they refer occur in the opposite order. Diagnostic learning is only possible because humans can use symbolic descriptions of causal situations that permit them to decouple the order of learning events from the order of real events. It is likely that a great deal of attention is required to separate learning order from causal order. Unless the instructed causal models are very plausible and accessible to participants, there may be great danger of falling back on the default assumption that the order of the experienced events follows the order of events in the world (i.e., predictive learning).

All these reasons lead us to the prediction that in complex conditions an *increase* of the blocking effect in diagnostic learning is predicted. Interestingly, causal-model theory predicts opposite effects on cue competition for predictive and diagnostic learning, less blocking in predictive but more blocking in diagnostic learning when the task is complex. Whenever the causal model is not salient or is unclear, we predict a tendency to fall back on the default assumption that causes precede their effects (i.e., predictive learning).

Distinguishing between competence and performance in causal learning aids the explanation of the inconsistent experimental evidence regarding sensitivity to causal directionality. A general pattern seems to be that participants tend to be capable of differentiating between predictive and diagnostic learning in relatively clear situations with few cues and outcomes in which intuitively plausible causal cover stories were provided (see Cobos et al., 2002; De Houwer & Beckers, 2002; Tangen & Allan, 2004; Tangen et al., 2005, for similar hypotheses).

By contrast, studies failing to reveal sensitivity to causal directionality tended to be more complex. Moreover, they often presented implausible or unclear causal cover stories (see also Waldmann, 2000; Waldmann & Holyoak, 1997). For example, Price and Yates (1995, Experiment 4) used cover stories that mentioned indicator lights or switches that were probabilistically related to the power output level of a nuclear power plant. It is possible that participants found it implausible that lights indicating nuclear power output level would give this information only probabilistically (see also Cobos et al., 2002).

Some experiments also present learning data that contradict the instructed causal model. For example, some of the experiments of Cobos et al. (2002) presented diseases with disjunctive effects (a disease could produce Symptoms D and E or Symptom F, but no other combination of these three symptoms). This pattern is clearly inconsistent with common-cause models, which would predict three conditionally independent effects (see also Waldmann, 2000; Waldmann & Holyoak, 1997). It is therefore possible that participants ignored the causal cover story, and fell back to predictive learning, the default learning procedure. In their Experiment 4, Cobos et al. used an overshadowing paradigm that avoided this problem. Unlike in Waldmann (2001) and Tangen and Allan (2004), who also used different types of overshadowing tasks with results supporting causal-model theory, Cobos and colleagues did not find a significant difference between predictive and diagnostic learning. The large number of cues and outcomes (nine cues, six outcomes) may be one reason for this failure. Additionally, Tangen and Allan speculated that the large number of trials may have been a contributing factor. Cobos et al. trained participants with maximally 210 trials (means over 64 trials), which according to the results of Tangen and Allan's experiments often lead participants to abandon the causal interpretation of cues and outcomes and treat the task as a simple associative contingency learning task.

The following three experiments represent different attempts in the direction of exploring boundary conditions for causal learning. In general, it is expected that human learners are sensitive to typical aspects of causal relations, such as their inherent directionality, in learning situations that place relatively low demands on processing capacity and clearly convey the underlying causal structure the learning events are referring to. Thus, in these conditions we expect to find evidence for people's competence to correctly acquire knowledge about causal models. Violations of these constraints, however, may limit learners' competency and foster the tendency to simplify the learning situation, for example by imposing a predictive-learning frame on the learning events or by ignoring relevant additional information, such as potential confounds.

Overview of Experiments

We have already pointed out that absence of blocking in predictive learning or presence of blocking in diagnostic learning might also be generated by prior domain assumptions (De Houwer, 2002; Waldmann, 2000, 2001). In this case different sizes of cue competition effects may be normative, so they do not count as evidence for performance effects. To be able to isolate performance limiting factors from such normative effects, in all three experiments cover stories are used that present fictitious devices. Since the devices are introduced as artifacts that were constructed to study learning, people should be less affected by prior assumptions about domains. Indeed, our previous experiments with artificial devices demonstrated that people had no problems grasping the intended causal model (see Waldmann, 2000, 2001).

Experiment 1 used the standard blocking paradigm with a predictive and a diagnostic cover story (see also Waldmann, 2000; Waldmann & Holyoak, 1992). Two additional factors were manipulated: The plausibility and tangibility of the causal model was varied by comparing a standard task in which the states of a causal device were verbally described on a computer screen with a task in which these states could directly be observed on a real device. Furthermore, information processing load was manipulated by means of a double task paradigm ("tone counting"). It is expected that sensitivity to causal directionality should display itself clearer in the conditions without the secondary task load and in the conditions in which the causal scenario was real.

Experiment 2 extends the range of cue competition paradigms by testing our hypotheses with an overshadowing paradigm. As in Waldmann (2000, 2001), we used fictitious devices with switches and lights as causes and effects. One goal of this study was to investigate whether the standard causal-model effect can also be shown with more complex devices with six cues and four outcomes. In the two standard conditions, a predictive version of the task was compared with a diagnostic version. A novel third condition presented a device in which participants had to learn both predictive and diagnostic relations within the same device. We expected this task to be considerably more complex than the standard tasks, and thus we were able to test our prediction that learning should suffer as a result of complexity.

Finally, Experiment 3 used a more complex blocking paradigm with diagnostic cover stories. Two slightly different phrasings of the cover stories were compared, which varied the amount of detail that was given in the introduction of the causal model. It is expected that the predicted absence of blocking after diagnostic learning depends on how the causal model is introduced in the initial instruction.

EXPERIMENT 1

The goal of this experiment was to test whether a manipulation of the information processing load and of the tangibility of the causal model affect sensitivity to causal directionality in a blocking paradigm. The participants' task was to learn about an artificial device, a black box that had colored lights on both sides (see also Waldmann, 2000, Experiment 1). One side represented the cause side, the other side the effect side. Each lamp on the cause side was linked to a button that was placed directly below it. Participants were told that pressing the button on the cause side caused the indicator light above the button to be on. In the predictive condition, participants received information about the states of several indicator lights on the cause side and had to predict whether pressing the respective button additionally caused the invisible light on the other side, the effect side, also to be on. After participants made their predictions, feedback was given about the presence or absence of the effect. In Phase 1, participants learned that one light on the front side caused the light on the back side also to be on (predictive cue). In contrast, a second light was nonpredictive (uncorrelated cue). In Phase 2, a third indicator light (redundant cue) was redundantly paired with the predictive cue. After each learning phase, participants were requested to rate the predictiveness of each individual light.

Associative theories as well as causal-model theory predict a blocking effect in this condition. According to the latter theory, two factors contribute to the predicted uncertainty about the redundant cue. First, no information is available about the influence of the redundant cue in the absence of the predictive cue since they are always presented together. Thus, it is impossible to compute the contingency between the redundant cue and the effect in the absence of the predictive cue. Second, the predictive cue deterministically causes the effect, which creates a ceiling situation when the predictive and redundant cues are present together. This makes it impossible for the new cue to display its potential causal power above the already established influence of the predictive cue.

For the diagnostic condition, the device was turned around. Two more lights were uncovered on the effect side, whereas on the cause side all but one light was covered. Now participants watched several effect lights that were potentially linked to the cause light on the invisible back side. Their task was to judge, on the basis of the state of the observed effect lights, whether the experimenter had pressed the button of this single cause light or not. Cues and outcomes were identical in the predictive and diagnostic conditions. Participants first saw the state of the effect lights, then made a judgment, and finally received information about the state of the cause light that indicated whether the experimenter had pressed the attached button or not. Again, participants first learned that one effect light (predictive cue) was turned on by the cause light on the invisible back side, whereas a second light was not causally affected by this light (uncorrelated cue). Then in Phase 2, a third light was uncovered. Now participants observed that two lights, the predictive and the redundant cue, were on whenever the cause light also was on. The same rating questions were used as in the predictive conditions. Because identical cues, outcomes, and test questions were used in the predictive and diagnostic conditions, associative theories again predict blocking of the redundant cue. By contrast, causal-model theory predicts that both lights should be seen as equally valid indicators of the cause light. Assessing contingencies in the cause–effect direction reveals that the cause light deterministically affects both the first and the third light.

To investigate performance limitations, two additional factors were manipulated in Experiment 1, tangibility and processing load. The tangibility and transparency of the underlying causal model was varied by comparing a condition in which participants learned about the causal relations by observing a *real* black box with buttons and lights with the standard case in experimental research with humans, verbal descriptions of the states of cues and outcomes on a computer screen. In general, computer presentations require a leap of faith on the part of participants. They read instructions about the causal meaning of the observed events but are certainly aware of the fact that the observed trials are just part of a computer program, and not the outcome of the causal mechanism they are supposedly learning about. In such conditions it may be tempting to forget about the causal cover stories. By contrast, observing a real mechanism should convince participants that the causal structure they are learning about is real. It should also serve as a constant reminder of the structure of the underlying causal model. Thus, it was expected that sensitivity to causal directionality should be particularly strong in the conditions in which participants learned about the real device.

To manipulate information processing load, a secondary tone-counting task was used in which participants were requested to count the occurrences of a high-pitched tone in a sequence of high- and low-pitched tones while learning about the device. It was expected that the effects predicted by causal-model theory should diminish in the conditions with tone counting. Both conditions, the predictive and the diagnostic conditions, should be affected. With additional complexity, participants are expected to prove reluctant to take cofactors into account in predictive learning, which, according to causal-model theory, underlies the blocking effect. Thus, *less* blocking is expected in this condition (see also De Houwer & Beckers, 2003). The diagnostic condition is also expected to be affected, with the increased processing load making it harder for participants to form and update a causal model that is directed opposite to the input order. These difficulties should increase the tendency to impose a predictive frame on the learning items, which should display itself in an *increased* tendency to block the redundant cue.

It is unclear whether associative theories would predict any effects of the secondary task. Given that cue competition is viewed as part of a basic, fundamental learning mechanism, it may be reasonable to expect blocking in both conditions, especially when processing demands prevent learners from using nonassociative strategies (see Price & Yates, 1995).

Method

Participants and Design. One hundred twenty-eight students from the University of Munich, Germany, participated in this experiment. Sixteen participants were randomly assigned to each cell, which was created by crossing the factors learning condition (predictive vs. diagnostic learning), tangibility of the causal model (real vs. described box), and processing load (tone counting vs. no tone counting).

Procedure and Materials. Prior to the learning task, participants in all conditions received written instructions (in German). In the *predictive-learning condition*, the instruction stated that the task would be to learn about causal relations. Participants read that they should imagine a box with three 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. Buttons were attached to the lamps on the front side. Pressing these buttons switched on the corresponding indicator light. The task was to learn to predict whether the button 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*, similar instructions were given. The only difference was that no buttons for the three 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 contained a button 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 they were expected to judge whether, as a consequence of the presumed actions of the experimenter, the indicator light of the cause was also lit.

The ensuing learning phase presented *identical learning trials* to participants in all conditions. The only difference was that the trials were either presented on a real box, or they were verbally described on a computer screen. In the computer version, participants were told that they were going to receive information about the states of three 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 might also be shown that indicate that the current state of the lamp could not 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 to 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 it would be useful to memorize the positions of the colors.

The learning trials showed a screen with the header "front side" above the three capitalized color names blue, yellow, and green next to each other on one line. In Phase 1, only information about two lights, the predictive and the uncorrelated cue, was given, and the other light was marked with question marks (e.g., "BLUE light: ON YELLOW light: OFF GREEN light: ????"). Thus, this example indicates that the blue light on the left side is on, the yellow light off, and the current state of the green light is unknown. The correct answer was to say "yes" when the predictive light (i.e., blue in this example) was on and "no" when it was off. The uncorrelated (yellow) light also varied between on and off, and the correct answer was always "no." This light was on only when the predictive light was off. Thus, there were three patterns in Phase 1 (correct responses in parentheses): on-off-???? ("yes"), off-off-???? ("no"), off-on-???? ("no"). After each judgment, the experimenter hit a key that displayed a screen with the feedback. Each pattern was presented eight times in random order. 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"). For half of the participants, the colors green and blue were exchanged. Thus, for these participants the green light was presented on the left side as the predictive cue.

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 rate how predictive each light individually was for the state of the light on the back side of the box. The sequence of colored lights followed the left-to-right sequence in the learning phase.

Before Phase 2 started, participants read further instructions stating that they were going to receive information about the uncovered third light. In the subsequent learning phase, three different trial types were presented again. These trials were identical with respect to the first two lights, the predictive and the uncorrelated cue.

The only difference was that now information about the third light (e.g., green) was also given. This light, the redundant cue, was always on when the predictive light was on (e.g., "BLUE light: ON YELLOW light: OFF GREEN light: ON") and off when the predictive light was off. Again, each pattern was presented eight times before the final ratings were collected.

The conditions in which a real box was used were run as closely as possible to the computer version. As a learning device, we built a box that was painted black and was 52 cm long, 31 cm wide, and 15.2 cm high. In the predictive condition, participants saw only the front side, and the back side was shielded by a black screen. The box was covered until after the initial instructions had been delivered. Then, participants in the predictive condition saw two (Phase 1) or three (Phase 2) colored lights (size 2.2 × 3.3 cm) that were connected to metal buttons. The buttons were placed directly below the indicator lights. Corresponding to the question marks in the verbal condition, the third light in Phase 1 was covered by a black metal plate so that the light underneath the plate was invisible. During the learning phase, the experimenter pressed the buttons on the cause side, which turned on the respective indicator lights. The pattern and sequence of buttonpresses corresponded to the trials in the verbal condition. Again, participants' task was to judge whether the light on the back side of the box was also on or off. After each of the initial trials in Phase 1, the shield was removed so that participants could check the state of the white outcome light on the back side. Later only verbal feedback was given. As in the verbal condition, participants saw each trial type eight times in each learning phase in random order. The colors of the predictive and redundant lights were also counterbalanced.

For the diagnostic condition, the box was turned around so that participants faced three effect lights. (The two additional lights visible in this condition were covered in the predictive condition, in which only one effect light was presented.) Otherwise, the cues, outcomes, trial types, and number of randomized trials were similar, as in the predictive condition. Participants observed the state of two (Phase 1) or three (Phase 2) lights on the visible front side. The only difference was that these lights did not have a visibly attached button. Instead the task was to judge whether the experimenter had pressed the button on the back side of the box, the cause side, which was attached to the white indicator light. Again a screen prevented participants from seeing the experimenter pressing the button. Instead they had to learn to use the visible effect lights as diagnostic cues. After each of the initial trials, the shield was removed after the judgment so that participants could see whether the experimenter had pressed the button by checking the state of the indicator light. Then the experimenter reset all the lights to being off. As in the predictive context, participants learned first that in Phase 1 one effect light (predictive cue) was perfectly correlated with the state of the cause light, whereas a second light was uncorrelated. In Phase 2, a third, redundant effect light was constantly paired with the predictive light. Now participants learned that the cause light was always on when both effect lights were on. After each phase, ratings were collected using the same instructions as in the verbal conditions.

The third factor manipulated the processing load. Half of the participants were confronted with an additional tone-counting task. During the learning phases, these participants heard a sequence of two different tones (1000 or 2500 Hz) in random order with intervals randomly varying between 3 and 7 sec. Each tone was presented for 0.5 sec. The task was to count the occurrences of the high-pitched tone.

Results and Discussion

The predictive cue received uniformly high ratings ($M = 96.5$) after Phase 1 training, which is a precondition for a potential blocking effect. These ratings did not differ across conditions.[2] By contrast, the uncorrelated cue received mean ratings of 6.68. Figure 1 shows the mean ratings for the predictive (P cue), redundant (R cue), and uncorrelated (U cue) cues, which were obtained after Phase 2. A 2 (predictive vs. diagnostic learning) × 2 (real box vs. described box) × 2 (tone counting vs. no tone counting) × 2 (predictive vs. redundant cue) analysis of variance (ANOVA) with the latter factor constituting a repeated measurement factor revealed a significant blocking effect (predictive vs. redundant cue) [$F(1,120) = 95.4$, $MS_e = 219.4$, $p < .001$] that interacted with learning condition (predictive vs. diagnostic learning) [$F(1,120) = 25.4$, $MS_e = 219.4$, $p < .001$] and the factor tangibility (real vs. described box) [$F(1,120) = 19.1$, $MS_e = 219.4$, $p < .001$]. Furthermore, the three-way interactions between the factors blocking, learning condition, and tangibility [$F(1,120) = 4.02$, $MS_e = 219.4$, $p = .047$], and between the factors blocking, learning condition, and processing load (tone counting vs. no tone counting) [$F(1,120) = 6.30$, $MS_e = 219.4$, $p = .013$], were significant. The four-way interaction was not significant.

To test the specific predictions of causal-model theory, we conducted a limited number of more specific analyses. A first analysis focused on the conditions *without tone counting*, which corresponded most closely to previous experiments (e.g., Waldmann, 2000). In this condition, behavior consistent with the normative predictions of causal-model theory should be observed in the condition with the real box. Thus, it is expected that blocking occurs only in the predictive condition and not in the diagnostic conditions, and that there is an interaction between the blocking effect and the learning condition. Moreover, it is expected that lowering tangibility by presenting the trials on a computer screen should weaken the predicted difference between predictive and diagnostic learning relative to the condition with the real box.

A 2 (predictive vs. diagnostic learning) × 2 (real box vs. described box) × 2 (predictive vs. redundant cue) ANOVA revealed a significant blocking effect [$F(1,60) = 69.09$, $MS_e = 158.8$, $p < .001$] that was moderated by two-way interactions with the factors learning condition [$F(1,60) = 39.4$, $MS_e = 158.8$, $p < .001$] and tangibility [$F(1,60) = 8.07$, $MS_e = 158.8$, $p = .006$]. Most importantly, the three-way interaction also proved significant [$F(1,60) = 4.88$, $MS_e = 158.8$, $p = .031$]. These results replicate previous findings (Waldmann, 2000; Waldmann & Holyoak, 1992) in that they show that participants were sensitive to causal directionality. Whereas the ratings between the predictive and the redundant cue clearly differed in the predictive conditions [$F(1,30) = 73.7$, $MS_e = 229.2$, $p < .001$], these cues failed to be significant in the diagnostic conditions [$F(1,30) = 3.7$, $MS_e = 88.4$, $p = .063$]. The latter result also rules out a possible alternative explanation of a blocking effect that attributes differential ratings to the unequal number of presentations of the predictive and redundant cue. If this

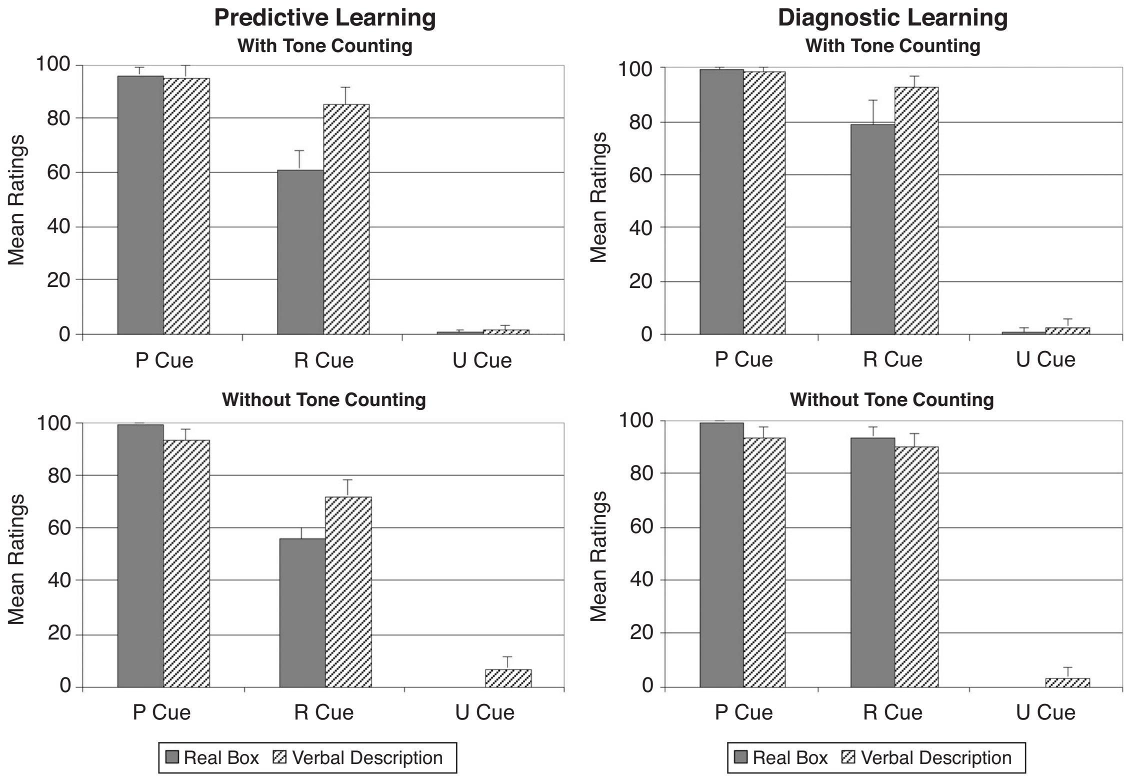

Figure 1. Mean "predictiveness" ratings (after Phase 2) for the predictive (P), redundant (R), and uncorrelated (U) cues in the predictive- (left) or diagnostic- (right) learning conditions. Learning trials were presented with or without a secondary tone-counting task either on a real box with buttons and lights, or verbally on a computer screen. The error bars represent standard errors of the mean calculated for each condition separately.

was the crucial factor, no interaction with learning condition should be observed (see also Waldmann, 2000).

In the predictive condition, the blocking effect was moderated by a significant interaction with the factor tangibility [$F(1,30) = 8.84$, $MS_e = 229.2$, $p = .006$]. As predicted, the blocking effect was more pronounced in the condition in which participants learned about the causal relations with a real box as opposed to a symbolic representation of learning trials on a computer screen. In the diagnostic conditions (without tone counting), tangibility did not have a detectable influence ($F < 1$, for the interaction).

The difference between the predictive and the diagnostic condition is also evident in the ratings of the redundant cues. These cues received higher ratings in the diagnostic than in the predictive condition with the real box [$F(1,30) = 46.8$, $MS_e = 244.5$, $p < .001$]. In contrast, there was a considerably weaker effect in the same direction in the condition with the computer presentation [$F(1,30) = 5.33$, $MS_e = 493.1$, $p = .028$].

The next analyses focused on the conditions *with tone counting*. Tone counting is also predicted to weaken the differences between predictive and diagnostic learning with respect to blocking. The analysis in which the predictive conditions were compared with the diagnostic condition again showed a significant blocking effect [$F(1,60) = 35.6$, $MS_e = 279.9$, $p < .001$] that was moderated by a two-way interaction with the factor tangibility [$F(1,60) = 11.1$, $MS_e = 270.9$, $p = .001$]. However, the two-way interaction between the blocking effect and learning condition [$F(1,60) = 2.51$, $MS_e = 279.9$, $p = .12$] and the three-way interactions ($F < 1$) both failed to be significant. The main reason for this new pattern is the fact that in this condition the redundant cue was rated significantly lower than the predictive cue not only in the predictive condition [$F(1,30) = 36.8$, $MS_e = 217.1$, $p < .001$] but also in the diagnostic condition [$F(1,30) = 7.8$, $MS_e = 342.8$, $p = .009$]. It is true that the effect in the diagnostic condition is clearly descriptively smaller than in the predictive condition, which indicates sensitivity for causal directionality also within this group of participants. But apparently the higher processing load had a detrimental effect on participants' ability to form causal models with learning input that was directed in the diagnostic direction. As in the conditions without tone counting, the blocking effect interacted with the factor tangibility in the predictive condition [$F(1,30) = 11.2$, $MS_e = 217.1$, $p = .002$], with a stronger blocking effect in the condition with the real box than in the context in which the trials were only verbally described on a computer screen. In contrast, the diagnostic condition was not affected by tangibility [$F(1,30) = 2.52$, $MS_e = 342.8$, $p = .12$, for the interaction].

Unlike in the condition without tone counting, there was no significant difference in the ratings of the redundant cues, either in the condition with the real box [$F(1,30) = 2.57$, $MS_e = 951.6$, $p = .12$] or in the condition with the computer presentation [$F(1,30) = 1.18$, $MS_e = 396.5$, $p = .29$], in the conditions with tone counting. This can be seen as evidence for the impact of the higher processing load caused by the tone-counting task.

In the final analyses, we checked the effect of tone counting. Generally, tone counting had a smaller effect on blocking than tangibility in the present experiment. However, in the condition with a real box, a significant three-way interaction between causal condition, the processing load factor, and the blocking effect [$F(1,60) = 4.03$, $MS_e = 272.8$, $p = .049$] was observed. As predicted, the blocking effect was smaller in the predictive condition when a tone-counting task added complexity to the task than without tone counting. The blocking effect was generally smaller in the diagnostic condition than in the predictive condition, which demonstrates sensitivity to causal directionality, but it can also be seen that the tendency to give lower ratings to the redundant cue than to the predictive cue (i.e., blocking) somewhat increases in the condition with tone counting. In the verbal condition, the three-way interaction did not prove significant, however [$F(1,60) = 2.28$, $MS_e = 165.9$, $p = .14$].

In summary, the results of the experiment showed once again that participants attempted to differentiate between diagnostic and predictive learning (see also Tangen & Allan, 2004; Waldmann, 2000; Waldmann & Holyoak, 1992). These findings are at odds with associative learning theories, which generally predict identical learning with identical cues and outcomes. However, the results also show that this competence may be corrupted by conditions that affect the salience of the causal relations as well as by the difficulty of the task. In the predictive conditions, blocking was more pronounced in the conditions in which participants were confronted with a real causal device as opposed to verbal descriptions of the learning trials on a computer screen. Apparently, the necessity to hold cofactors constant becomes more apparent in a situation in which the converging causal influences could be imagined more vividly. Processing load also had a small effect on learning. The normative predictions of causal-model theory were borne out in the conditions without tone counting (especially when participants learned with the real box). In the conditions with increased processing load, the blocking effect slightly decreased in the predictive and increased in the diagnostic conditions relative to the corresponding conditions without tone counting. This pattern of opposite influences of performance limiting factors on predictive versus diagnostic learning can be predicted by an extension of causal-model theory, whereas it appears hard to reconcile these findings with associationist theories.

The decrease of the blocking effect in the predictive-learning conditions is consistent with the study of De Houwer and Beckers (2003). One interesting difference is that De Houwer and Beckers found an effect of the secondary task only when it was presented during both learning and test phases, whereas we found an effect, albeit a small one, in a task in which participants were confronted with the secondary task only during the learning phase. Possibly our task was more sensitive to differences because we compared learning with and without a secondary task, whereas De Houwer and Beckers used two differently difficult versions of the tone-counting task.

One interesting difference between our theory of the effect of processing load and the theory of De Houwer and Beckers (2003) concerns the underlying mechanism. Whereas our theory claims that people should tend to fall back to using simple unconditional contingencies under conditions of heavy load, De Houwer and Beckers predict that a secondary task should prevent learners from making the right inferences about the causal status of the redundant cue, and therefore lead to uncertainty. In our view, the results of the present experiment favor our theory. We used deterministic relations, which according to De Houwer and Beckers should already create uncertainty about the causal status of the redundant cue in regular learning situations without a secondary task. Nevertheless we found an effect of tone counting and tangibility in our deterministic scenario. Moreover, the ratings of the redundant cue were on an intermediate level (far from the ones for the uncorrelated cue) in the conditions without tone counting and the real box, and moved toward the predictive cue in the conditions with tone counting and verbal descriptions. Thus, there was a tendency to move from partial blocking toward absence of blocking with increasing processing load and decreasing tangibility, rather than a tendency to move from absence of blocking toward partial blocking (see note 1), which is consistent with our theory. More experiments are certainly needed to conclusively resolve this issue.

In summary, the results of the experiment demonstrate that people attempt to learn correctly about causal models if possible. However, they often fail in situations with increased processing load and decreased tangibility of the causal model.

EXPERIMENT 2

Experiment 1 has shown that both predictive and diagnostic learning suffer when there is a secondary task, or when the tangibility of the causal domain is low. However, the detrimental effect on diagnostic learning (i.e., blocking) was relatively small. This is peculiar because other research groups have found much larger blocking effects in diagnostic-learning tasks (e.g., Cobos et al., 2002; Price & Yates, 1995). There are several possible reasons for these divergences. Typically more cues, outcomes, and learning trials have been used, all of which may lower the plausibility of the causal nature of the task and lead participants into an associative learning mode (see De Houwer & Beckers, 2002; Tangen & Allan, 2004). Differences in domain characteristics may also be a factor. Devices with switches and lights are domains in which the underlying causal structure is very salient in comparison with more opaque domains, such as diseases (see Waldmann, 2001).

The goal of Experiment 2 was to test predictive and diagnostic learning in conditions that are more demanding while sticking to artificial devices. The most important novel contribution of this experiment is that we included a *mixed-learning* condition in which learners were confronted with a device in which causes and effects were both presented as cues. Thus, participants had to learn both predictive and diagnostic relations within a single device. We used a device with two halves. On one half, participants learned about three cause cues and two effect outcomes, whereas on the other half, three effect cues and two cause outcomes were presented (Figure 2). The two halves were presented consecutively prior to the test phase with the predictiveness rating questions for both halves.

We expected that this condition would be particularly hard because participants learn about identical statistical relations for the cue–outcome relations on both sides, which nevertheless should be processed differently, depending on the underlying causal model. Thus, identical statistical relations between cues and outcomes had to be assigned to different causal models (common cause vs. common effect), which makes it necessary to mentally reorganize the learning input in a way that predictive and diagnostic relations are correctly represented within a single coherent causal model of the device. To adequately learn the predictive relations, confounded causes have to be taken into account, whereas correctly learning the diagnostic relations implies a mapping from cues to effects and outcomes to causes and the assessment of unconditional contingencies in the cause–effect direction. We expected that in this mixed condition participants might tend to fall back on the more natural predictive-learning mode and show a clear cue competition effect for both predictive and diagnostic relations. For the predictive relations, the effect might be smaller due to the heightened complexity.

To compare performance in the mixed condition with that in our standard conditions, we also ran a pure predictive-learning condition with six cause cues and four effect outcomes that were also arranged on two halves of the device. In the corresponding pure diagnostic-learning condition, we presented six effect cues and four cause outcomes in a similar fashion.

To test our hypotheses with a paradigm different from blocking, we chose an overshadowing task (see also Waldmann, 2001). In all conditions a single *predictive* cue on each side was deterministically connected with one outcome. Trials showing the single predictive cues were intermixed within each side with trials that showed two redundant cues followed by a second outcome. Associative learning theories (e.g., Rescorla & Wagner, 1972) predict an overshadowing effect, which should display itself in lower ratings of the redundant cues compared with the single cues. Causal-model theory (see Waldmann, 2001) predicts overshadowing in the predictive conditions but not in the diagnostic conditions, at least in the pure conditions. In the predictive conditions, only the single cause is presented by itself as a deterministic cause, whereas none of the redundant causes are presented in the absence of the potential alternative cause. Thus participants should prove uncertain about the status of the redundant causes, which both are perfectly confounded. In contrast, in the diagnostic conditions, all effects are deterministically generated by a single unique or a single common cause. Given that alternative hidden causes seem implausible in such devices, participants should tend to give similarly high diagnostic ratings for single and redundant effects.

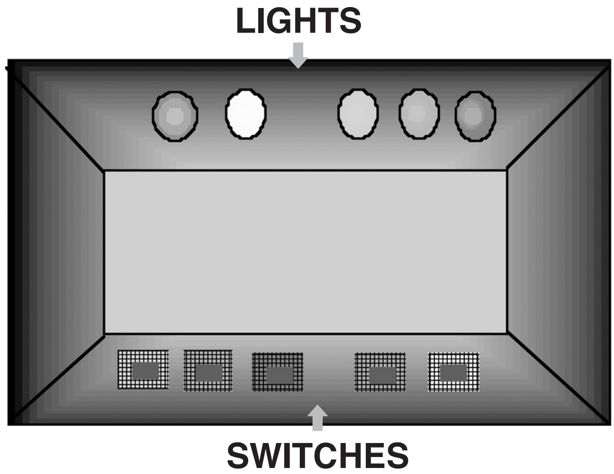

Figure 2. Example of a device used in the mixed condition of Experiment 2. Switches represent causes, and lights effects. In this particular counterbalancing condition, the left half was used for predictive learning of the relation between three causes and two effects (cues on the bottom), and the right half for diagnostic learning of three effects and two causes (cues on the top). The two other conditions (pure predictive or diagnostic) used devices that duplicated either the left or the right half of this device.

Method

Participants and Design. Eighty-four undergraduate psychology students from the University of California, Los Angeles, participated in exchange for course credit. Twenty-eight of those students participated in the pure predictive, 28 in the pure diagnostic, and 28 participated in the mixed-learning condition.

Materials and Procedure

The experiment was presented to each participant on a personal computer. Each condition was based on the overshadowing paradigm (see also Waldmann, 2001). For all conditions, participants were given the same cover story, stating that we were interested in investigating how people learn to operate devices such as VCRs and computers. Participants were then told that we had constructed an

Table 1
Design and Trial Types of Experiment 2

Learning Condition	Trial Type
Pure predictive	Cause A→Effect 1
	Causes B + C→Effect 2
	Cause D→Effect 3
	Causes E + F→Effect 4
Pure diagnostic	Effect A→Cause 1
	Effects B + C→Cause 2
	Effect D→Cause 3
	Effects E + F→Cause 4
Mixed predictive and mixed diagnostic	Cause A→Effect 1
	Causes B + C→Effect 2
	Effect D→Cause 3
	Effects E + F→Cause 4

Note—The cues are listed on the left side and the outcomes on the right side of the arrows, which represent learning order. The letters of the cues and the numbers of the outcomes denote different colors of switches (causes) and lights (effects).

artificial device that we would present to them on the computer screen, and that the device has both lights and switches. We pointed out that lights are only turned on by switches to discourage the assumption of hidden causes. Finally, participants were told that it was their task to learn the relationships between those switches and lights, and that we would check at the end of the experiment if they knew how to operate the device. In the initial instructions, we showed participants a picture depicting the full device with six cues and four outcomes (see Figure 2 for an example from the mixed condition).

In each condition, participants then entered a learning phase in which they learned the particular relationships between the cues and outcomes. In all cases, the causes were switches and the effects were lights. Each device in each condition had six cues and four outcomes. All presented cue–outcome relations were deterministic; that is, the cue predicted the outcome in 100% of the trials. Also in all conditions the switches (i.e., causes) were on one side of the device (top or bottom) and the lights (i.e., effects) on the other side (see Figure 2 for an example). The mapping of causes and effects to either of these sides was counterbalanced across participants. We decided to present the device this way to simplify the device and the assignment of the causal status to cues and outcomes.

Table 1 displays the design and trial types for all three conditions. In the *mixed-learning* condition, the device was divided in half so that half of the device contained three causes and two effects (for predictive learning), and the other half contained two causes and three effects (for diagnostic learning) (see Figure 2). While participants were learning about one half of the device (left or right), the other half was covered. The participants were told that this would help them concentrate by eliminating distraction. Whether participants started with predictive or diagnostic learning was counterbalanced.

In the predictive-learning part of the task, participants received two blocks with 15 trials each presented in random order. Five trials were shown in which participants observed the presence of a single predictive cause (switch), which deterministically led to the presence of the corresponding effect (light). Five trials presented two causes (switches) simultaneously (redundant cues) paired with the second effect, and five trials presented cases in which all causes and consequently all effects were absent. Each trial consisted of a question, a response, and then feedback as to whether the answer was right or wrong and what the right answer was. First we showed participants which switch(es) were on and then asked them, "Which light is on?" Participants were instructed to press the appropriate key to indicate their response. To help them choose the right key, we gave them a printed sheet that showed which light or switch corresponded to which key (e.g., "if you think the blue light is on, press B," etc.).

After the first learning phase, the half of the device that participants had just learned about was covered, and the other half was shown. If the first task was a predictive-learning task, the second task was a diagnostic-learning task with three effect cues (lights) and two cause outcomes (switches) (see Figure 2). Now participants received information about the presence or absence of lights (effect cues), and were asked, "Which switch is on?" As in the predictive-learning phase, there were two blocks of 15 trials in random order. Each block consisted of 5 trials in which one light was caused by one of the two switches, and two lights that were on together were caused by the second switch (Table 1). There were also five trials in which all lights and switches were off.

The learning trials were blocked according to the half of the device, and the position of cues and outcomes (bottom or top side) was counterbalanced across participants. The order of blocks was counterbalanced across participants. Also, whether a particular side of the device was for predicting or diagnosing was counterbalanced across participants. Thus, in one counterbalancing condition, participants might receive three cause cues on the lower side of the device and learn to predict two effect cues on the upper side first (predictive learning), and then they were presented with three effect cues on the upper side of the device, and learned to diagnose two cause cues on the lower side (diagnostic learning).

In the test phase, which followed the learning phase, the ratings for all six cues (three causes, three effects) were collected. In the diagnostic part of the test phase, participants were asked to rate how predictive each light (i.e., effect cue) was for a particular switch (i.e., cause outcome) being on, and, in the predictive part, how predictive each switch (i.e., cause cue) was for a particular light (i.e., effect outcome) being on. Thus, in the diagnostic part, the lights (effects) are treated as cues, and in the predictive part, the switches (causes) are treated as cues. Each half was presented separately with the other half being covered. For each test question, all but one cue was covered by a rectangular area that allowed learners to see only the current state of the cue that should be rated. Participants were not informed about the state of the switches or lights that were covered. Participants were instructed to press the "P" key if they wanted to answer "perfectly predictive," the "M" key if they wanted to answer "medium predictiveness," and the "N" key to answer "not predictive." We chose these three responses because previous research has shown that in deterministic situations, participants tend to spontaneously say that a causal relation is either present or absent, or that they do not know on the basis of what they saw. We counterbalanced the order in which the halves were presented across the learning phase and the rating phase so that we had every possible combination of orders represented (e.g., learn about the right half then the left, and then the rating questions start on the left; learn about right half then left half, and then the rating questions start on the right half, etc.). Because every cue had to be rated with respect to two outcomes on the respective half, overall each participant was presented with 12 test questions (6 on each half).

The procedure of the "pure" conditions closely mirrored that of the mixed condition. Basically the *pure predictive* condition duplicated the predictive-learning part of the mixed condition, and the *pure diagnostic* condition the diagnostic-learning part (see Table 1). Thus participants in the pure predictive condition were presented with a device that again was divided in half. During the learning and the final rating phase, one half was always covered before learning or testing proceeded to the other half. Each half showed three switches as cause cues on one side and two effect lights as outcomes on the other side. On each of the two halves, one switch (cause) predicted one effect, and two switches together the second effect. Learning instructions, task, and number of trials for each half were identical to the one in the predictive-learning part of the mixed condition. Again the test phase followed learning of the two halves, with virtually the same procedure as in the predictive part of the mixed condition.

In the *pure diagnostic* condition, participants were presented with six effect cues (lights) and four cause outcomes (switches) (see Table 1). This condition duplicated the diagnostic-learning part of the mixed condition. Thus, learning and testing were directed in the effect-cause direction on each half. On each half, one effect (light) was caused by one cause (switch) and the two remaining effects by the second cause. Otherwise the same procedure and the same counterbalancing of the position of switches and lights were used as in the other conditions.

Results and Discussion

We assigned a numerical value to each type of response for the rating phase (*not predictive* = 0, *medium predictiveness* = 1, and *perfectly predictive* = 2). A preliminary analysis showed that predictive and redundant cues were rated similarly within each learning condition. Therefore, we used the mean ratings for the average of the single cues and the mean ratings for the average of the redundant cues as dependent variables.

Figure 3 shows these mean ratings for the predictive and diagnostic conditions. As can be seen from the figure, all participants gave a rating of perfectly predictive for the single cue (in all conditions). Also, nearly all participants in the pure diagnostic condition rated the redundant cues as "perfectly predictive." In contrast, the ratings for the redundant cues in the pure predictive condition were nearly all at the level of "medium predictiveness" despite the fact that these participants received the same learning input as the ones in the diagnostic condition. All participants in all conditions rated the nonrelated cue–outcome pairs (not shown in Figure 3) as "not predictive" ($M = 0$). These relations were not further analyzed.

Thus, in the pure conditions participants showed assessments that support the normative predictions of causal-model theory. Whereas overshadowing was observed in the pure predictive condition, no such effect is seen in the pure diagnostic condition. The results look different in the mixed condition, though—both the predictive and diagnostic ratings center on "medium predictiveness."

The statistical analyses confirm this impression. We used an initial ANOVA for repeated measurements to show that the difference between the mixed-predictive and mixed-diagnostic redundant cue ratings was not significant [$F(1,27) = 1.85$, $MS_e = .02$, $p = .18$], which supports our prediction of overshadowing in the diagnostic phase during the complex task. The absence of a reliable difference between the redundant cue ratings reveals that participants in this condition apparently were not sensitive to the difference between the two learning conditions within this complex task, predictive and diagnostic learning. A 3 (pure predictive vs. pure diagnostic vs. mixed) × 2 (single cue vs. redundant cues) ANOVA with cue type as a within-subjects factor and the redundant cue averaged over all four measurements in all conditions yielded a highly significant interaction between learning condition and cue type [$F(2,81) = 82.6$, $MS_e = 0.04$, $p < .001$], a significant effect of cue type [$F(1,81) = 356.5$, $MS_e = 0.04$, $p < .001$], and a highly significant main effect for learning condition [$F(2,81) = 62.9$, $MS_e = 0.06$, $p < .001$].

The following planned comparisons provided a more direct test of our hypotheses. First we focused on the pure conditions, comparing the averaged redundant cues with the single predictive cues, which tests for the presence of an overshadowing effect. For these analyses, the four ratings of each of the pure conditions and the two

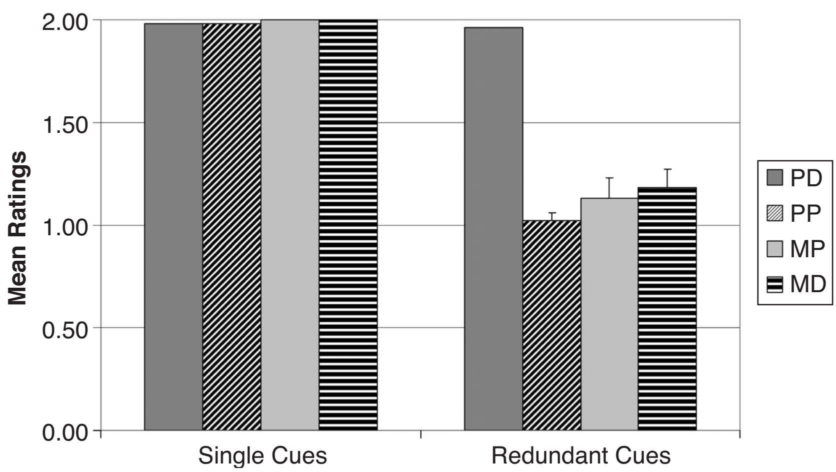

Figure 3. Mean "predictiveness" ratings for the averaged single and redundant cues in the conditions "pure diagnostic" (PD), "pure predictive" (PP), "mixed predictive" (MP), and "mixed diagnostic" (MD). The error bars represent standard errors of the mean (which are <.01 when invisible) calculated for each condition separately.

ratings of either the predictive or the diagnostic subtask of the mixed condition were averaged. These tests showed a significant overshadowing effect only in the predictive condition [$F(1,27) = 1{,}514.1$, $MS_e = 0.009$, $p < .001$], but not in the diagnostic condition [$F(1,27) = 1.86$, $MS_e = 0.005$, $p = .18$]. The ratings for the redundant cues were clearly different in the pure predictive and pure diagnostic conditions [$F(1,54) = 532.5$, $MS_e = 0.02$, $p < .001$]. This pattern is consonant with causal-model theory, which predicts overshadowing in the predictive but not in the diagnostic task.

In the mixed condition, however, there was a significant overshadowing effect for both the predictive-learning component [$F(1,27) = 74.1$, $MS_e = 0.14$, $p < .001$] and the diagnostic-learning component [$F(1,27) = 90.9$, $MS_e = 0.10$, $p < .001$]. The comparisons of the ratings for the redundant cues for the predictive-learning component showed no significant difference from the ratings in the pure predictive condition [$F(1,54) = 1.08$, $MS_e = 0.15$, $p = .30$], whereas there was a clear difference between the ratings of the redundant cues in the mixed diagnostic-learning versus the pure diagnostic-learning conditions [$F(1,54) = 68.9$, $MS_e = 0.12$, $p < .001$]. This pattern supports the hypothesis that diagnostic learning should suffer in a mixed-learning task. Apparently there was a tendency to fall back on predictive learning because the task became too complex.

In the present experiment, we did not see a significant decrease of blocking for the predictive component. However, descriptively there was a tendency in this direction. There were fewer "medium predictiveness" responses in the mixed predictive condition than in the pure predictive condition; perhaps the rating measures were not sensitive enough to pick up a reliable difference. It may also be the case that a decreased cue competition effect with additional processing load is more likely in two-phase blocking designs than in overshadowing tasks because in blocking designs the redundant cue is directly paired with a perfectly predictive cue.

One practically interesting result of this experiment is that people had difficulty learning about mixed devices that contained both cause–effect and effect–cause components. This is peculiar because real devices (e.g., VCRs) often have both switches (causes) and indicator lights (effects) on the front side. One important difference is that our devices, which were introduced as artificial devices designed for studying human learning, violated the one cause–one effect constraint typical of such devices. Usually a switch has a single function, and an indicator light signals one type of state. Also, in real-world learning we would test each switch individually to avoid the confounds inherent in cue competition paradigms. Thus, the learning device in our experiment is more difficult to learn than most real devices (which often are also hard to understand). We expect that people can learn about mixed devices when the causal relations are simplified or when the task and the instructions are presented in a way that makes learning easier. Thus, we do not claim that people are generally incapable of learning about mixed devices; rather, we believe that these devices are more difficult to learn than similarly complex pure devices.

EXPERIMENT 3

A further factor contributing to performance limitations may be the plausibility and clarity of the cover stories. As was pointed out in the introduction, recent failures to find sensitivity to causal directionality in diagnostic-learning tasks (i.e., absence of blocking with common-cause structures) may have been partly due to the fact that the chosen cover stories were implausible or the mapping of cues and outcomes to the roles of causes and effects was not clearly conveyed in the instructions (see Waldmann & Holyoak, 1997).

The present experiment explored the role of the phrasing of the cover stories in a diagnostic-learning task. Again we used a device with switches and lights as the learning task (similar to the one in Experiment 1). All participants received information about the trials on a computer screen. The crucial manipulation involved the cover stories. A similar cover story was used as in the diagnostic condition of Experiment 1. Thus, participants read about a box with lamps on the front side and a single lamp on the back side. In one condition, the *elaborate* condition, participants were instructed that occasionally a person presses a button attached to the indicator light on the invisible back side that causes this light to go on. The question was whether the pressing of the button additionally has an effect on the lamps on the visible front side. The participants' task would be to learn to infer the behavior of the lamp on the back side on the basis of the behavior of the lamps on the visible front side. Thus, the light on the back side was introduced as an indicator of a potential common cause of the states of the lights on the visible front side, which therefore were potential effects.

In the contrast condition, the *vague* condition, the instruction was slightly altered. Instead of mentioning a person occasionally pressing the causal button on the back side, participants just read that the lamp on the back side was a potential cause of the lighting up of the lamps on the front side. Again it was pointed out that the goal was to find out which lamps on front side were influenced by the lamp on the back side, and that the task would involve inferring the behavior of the lamp on the back side on the basis of the behavior of the lamps on the front side.

It was expected that participants in the vague condition would have greater difficulty forming a sensible causal model with a common-cause structure. Even though it was mentioned that the lamp on the back side was a potential cause, these instructions are, at least, incomplete. According to prior world knowledge, lamps by themselves are not causes. Typically they are effects (of switches) or indicators of intermediate causal events in a causal chain. Thus, unlike in the elaborate instructions, the cover story in the vague condition leaves open how the light can possibly be a cause of the lights on the front side. The opaqueness of the cover story may lead many participants to ignore the cover stories and resort to a simpler strategy in which the temporal order of the learning events corresponds to the causal order, thus treating the cues like causal indicators or noncausal predictors of the predicted event.

A causal scenario slightly more complex than that of Experiment 1, with four lamps on the front side, was used in this experiment (see also Waldmann, 2000, Experiment 1). In Phase 1, participants learned that the lamp on the back side was causally related to one light on the front side (predictive cue). The other lights were covered so that information about the potential causal influence on the remaining three lights was not available. In Phase 2, two pairs of lamps were alternately uncovered. Thus, in each trial participants only received information about two lights. Either participants saw the predictive light from Phase 1 along with a second light, the redundant cue, lighting up together, or the remaining two lights could be seen lighting up together (the informative cues). In both cases, the lighting up of the visible pair of lights indicated that the causal light on the back side had been switched on. Thus, participants should learn that the light on the back side deterministically influenced all four lights on the front side, although there were no trials on which every light was visible. Eventually all lights should, according to causal-model theory, be viewed as equally valid effect indicators of the cause light by the learners, which indeed was demonstrated in Experiment 1 of Waldmann (2000).

This paradigm, which was adapted from studies by Chapman and Robbins (1990) and Williams, Sagness, and McPhee (1994), allows testing of the blocking effect by comparing ratings of cues that have been presented an *identical* number of trials within compounds (see also Waldmann, 2000). Full blocking of the redundant cue by the predictive cue should be indicated by a significant difference between the ratings of the redundant and the informative cues, which correspond to an overshadowing control condition (see also note 1). According to the Rescorla–Wagner (1972) rule, the predictive cue, in case it is trained to the asymptote in Phase 1, should completely block the redundantly paired cue in Phase 2. In contrast, the two informative cues in Phase 2 should rise at equal pace until the sum of their associative strengths fully predicts the outcome. Assuming equal salience and equal learning rates, the two informative cues should gain intermediate associative strength, whereas the redundant cue should be kept at a lower level by the predictive cue.

This task is more complex than the task in Experiment 1 not only because of the increase in the number of cues but also because it requires learners to continually keep track of currently visible and invisible lights. According to causal-model theory, the inference that all lights were always on when the cause was present requires the (plausible) background assumption that the causal influence of the common cause remains stable during the learning period so that the contingencies observed for the visible cues can be generalized to trials in which they were temporarily covered. It seems likely that participants would be more willing to make these inductive generalizations when they are confident about the structure of the underlying causal model.

Method

Participants and Design. Forty-eight students from the University of Munich, Germany, participated in this experiment. Half of this group was randomly assigned to the condition with the elaborate instruction, the other half to the condition with the vague instruction.

Procedure and Materials. Participants received written instructions (in German); these were summarized in the introduction to this experiment. In the condition with the *elaborate* instruction, the cause light on the back of the black box was introduced as being connected to a button that a person occasionally pressed. Pressing the button causes the corresponding indicator light to go on, and the question was to find out whether the pressing also causally influences the visible lights. The task was to infer the state of the light on the back side on the basis of the behavior of the lights on the front side. Participants were requested to say "yes" or "no" depending on whether they thought the light on the back side was on or off.

In the condition with the *vague* instruction, neither a button nor a person who manipulated the cause light was mentioned. The light on the back side was simply introduced as a potential cause for the lighting up of the lights on the front side. Otherwise the instructions were as in the elaborate condition.

Next, the learning trials on the computer started. All participants received *identical* learning trials with identical learning instructions. These instructions pointed out 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 stated that instead of this information, four question marks might also be shown indicating that the current state of the respective lamp could 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 to say "no" when it was presumably off. After their decision, they would be given immediate feedback. Then participants were alerted to memorize the positions of the lamps because later they would be asked about the different lamps.

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, and 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. After each decision, the experimenter hit one key that 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"). Six trials were presented to the participants in this phase (three "yes" trials, three "no" trials). For half of the participants, the red light on the left side and for the other half the blue light on the right side was the predictive cue.

After this learning phase, participants were requested to rate the predictiveness of the lights 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 predicted 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 states of the other two lights were masked with question marks. Two trial types consisted of pairing the predictive light from Phase 1 with a new redundant light. Either both lights were off or both lights were on (e.g., "RED: ON GREEN: ???? WHITE: ON BLUE: ????"). The other two trial types presented the two other lights either both on or both off (e.g., "RED: ???? GREEN: ON WHITE: ???? BLUE: ON"). Whenever the lights in either of the two trial types 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 lights that were uncovered in Phase 2 was counterbalanced. After the learning phase, participants again rated the predictiveness of the four lights. The order of the rating questions corresponded to the left-to-right sequence of the four lights in the particular counterbalancing condition to which the participant was assigned.

Results and Discussion

All participants gave the predictive cue the maximal rating of 100 after Phase 1. Figure 4 shows the mean ratings of the three cue types after Phase 2. The two informative cues were averaged. All cues received virtually identical mean ratings in the condition with the elaborate instructions, whereas clear differences can be seen in the contrasting condition with the vague instruction.

The strongest test of a full blocking effect involves a comparison between the redundant and the average of the informative cues (i.e., the overshadowing control). Both types of cues have been presented an equal number of times as parts of compounds. A 2 (redundant vs. informative cues) $\times$ 2 (elaborate vs. vague instructions) ANOVA with the first factor constituting a repeated measurement factor showed a significant difference between the two cues [$F(1,46) = 14.22$, $MS_e = 333.7$, $p < .001$] that interacted with the type of instruction [$F(1,46) = 4.39$, $MS_e = 333.7$, $p = .042$]. Whereas no significant difference between the two cue types was observed in the elaborate condition [$F(1,23) = 1.86$, $MS_e = 251.4$, $p = .19$], the ratings clearly differed in the vague condition [$F(1,23) = 13.8$, $MS_e = 416.1$, $p = .001$].

In summary, the condition with elaborate instructions replicates previous experiments (e.g., Waldmann, 2000, Experiment 1) and confirms, once again, causal-model theory's prediction that no blocking should be observed with effect cues that are equally influenced by a common cause. This finding is at odds with the predictions of associative learning theories. However, the condition with the vague instruction shows results that are virtually identical to the results typically seen for predictive versions of the task (see Waldmann, 2000, Experiment 1). As predicted, the vague and incomplete nature of the cover story in this condition seems to have weakened the tendency of participants to map the learning cues to the effect layer in a common-cause model. Not only did the task used in this experiment require reorderings of the learning events from their diagnostic effect–cause presentation order to the cause–effect order in the causal model; the difficulty of the task was furthermore exacerbated by the fact that periodically information about the cues was withheld from the participants. The cues were alternately hidden so that it was necessary to infer the states of hidden cues on the basis of previous learning trials. The conclusion that all cues were equally affected by the common cause relied on the assumption that the cause did not change its deterministic capacity between trials in which the effects were visible and trials in which they were not. It seems plausible that this additional inductive step made the task more demanding,

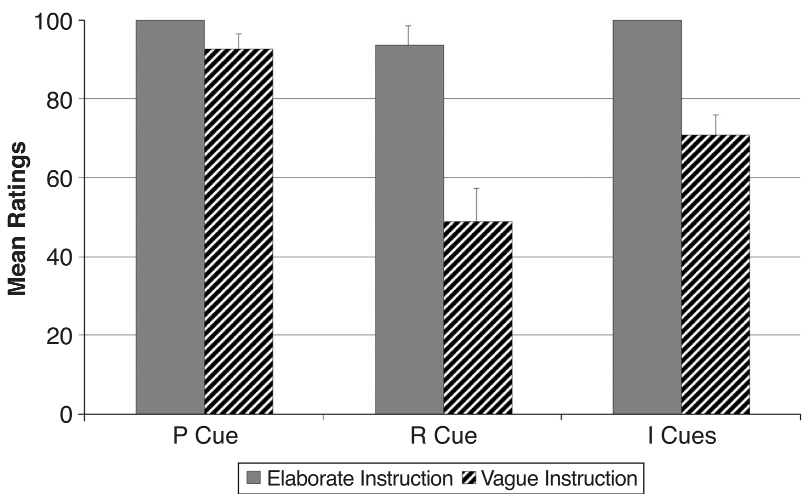

Figure 4. Mean "predictiveness" ratings after Phase 2 of diagnostic learning for the predictive (P), redundant (R), or average of the informative (I) cues in the conditions with elaborate or vague causal instructions. The error bars represent standard errors of the mean (which are <.01 when invisible) calculated for each condition separately.

which may have increased the tendency to ignore the common-cause instructions in the condition in which the instructions were more implausible and vague.[3]

GENERAL DISCUSSION

One goal of the experiments was to provide further evidence for causal-model theory's claim that human learners attempt to form representations of the world that honor physical characteristics of causality, such as the inherent directionality of cause–effect relations. The three experiments showed that in conditions with relatively little processing demands and with clearly conveyed causal models, most participants correctly learned different causal models. These findings replicate previous experiments (Waldmann, 2000, 2001; Waldmann & Holyoak, 1992) and provide further support for the claim that associative theories, being insensitive to the causal roles of cues and outcomes, are unable to model human learners' competence to acquire causal knowledge.

However, the experiments furthermore show that this competence does not display itself in all circumstances. In particular, when the task is complex or the causal structure is not salient or sufficiently transparent, participants tended to fall back on a modus that corresponded to predictive cause–effect learning even when the cues were meant to represent effects and the outcome a cause. In predictive learning, the temporal order of learning events matches the temporal order of causal events, whereas in diagnostic learning, there is a mismatch. Thus, this type of learning task requires a cross-mapping between cues and effects, and outcomes and causes, which is computationally demanding.

The results also show that performance factors not only jeopardize diagnostic learning but also affect predictive learning. With increased processing load and less tangibility of the causal structure, learners showed less blocking, which, according to causal-model theory, is a consequence of the reduced ability to control for the influence of cofactors (see also De Houwer & Beckers, 2003). In an overshadowing paradigm (Experiment 2), we did not find a significant decrease of cue competition in the mixed condition, however. We speculate that the direct pairing of a redundant cue with a predictive cue in the blocking paradigm is an important factor that may tempt participants to give similar ratings when the task is difficult.

These results raise the question of what relevance competence versus performance has for everyday learning. Cobos et al. (2002) acknowledged that normative learning might be seen in simple scenarios but argued that in more "realistic" domains people learn using associative learning mechanisms. We believe that this evaluation is inadequate. Very little is known about how people learn about causal relations in the real world. The tasks typically studied in the laboratory are certainly different from those in real-world learning contexts. People usually do not have to learn about complex causal models with nine cues and six outcomes within a brief learning period. It is more natural that we only learn about small fragments of causal models, and later integrate the fragments to more complex models. The present research shows that people try to form adequate representations of causal models in situations that do not transcend their capacity. In laboratory experiments, people are often forced to learn anyway, but in the real world they might select their learning environment in a way that matches their capacity.

Another unrealistic feature of many laboratory experiments is that in virtually all studies people learn about an artificial domain on a computer with stimuli that, as every participant knows, are generated by a programmer and not by the causal mechanisms mentioned in the instructions. This often leads participants to ignore the initial instructions, especially when there are many trials (see De Houwer & Beckers, 2002; Tangen & Allan, 2004, in press). In our research, we often let participants summarize the cover story prior to learning to make sure that the cues and outcomes are interpreted causally. In real causal contexts, this may not be necessary because the learning events normally remind people of their causal nature (see Experiment 1).

The results of the experiments raise the question of learning mechanisms that may account for the data pattern. It is clear that both associative theories and causal-model theory in its original version cannot account for the full pattern of results. One obvious alternative, suggested by Price and Yates (1995), is a two-process theory that embodies both an associative learning component and a rule-based component (see also Tangen & Allan, 2004). According to this theory, trial-by-trial on-line learning is managed by an associative learning mechanism that is directed from cues to outcomes. In parallel, event frequencies are stored that are handled by the rule-based component. Whenever the judgment task is directed from cues to outcomes, the associative weights are accessed, whereas other tasks (e.g., estimation of frequencies or of the strength of the outcome–cue relation) are processed by the rule-based component, which combines frequency information according to judgment rules. Price and Yates also speculated that rule-based processing is more likely with tasks that are not too demanding.

Although it is likely that several processes interact in learning, this particular two-process model (Price & Yates, 1995) explains the present results only in part. The results that conform to the predictions of causal-model theory may be accounted for by the rule-based component. However, it is less clear how the impact of the phrasing of the cover stories or the mode of presentation of learning trials would be integrated in this model. The greatest problems arise with the finding that cue competition seems to decrease in predictive learning when the complexity of the task is increased or when the causal structure is presented less saliently. If anything, these conditions seem to be harder than the contrast conditions, which, according to the two-process model, should increase the likelihood that an associative mechanism with its built-in mechanism of cue competition would take over.

In our view, the most promising general approach to the psychology of causality is the research strategy most linguists choose for the analysis of language (Chomsky, 1965). We should start with models that describe our competence before we deal with conditions that prevent people from displaying their competence. If we want to learn about our language faculty, then we should give participants the opportunity to show what they can do. Later we can study errors people make under different conditions. Similarly, it is more informative to explore what people can do when learning about causal relations under optimal conditions before we investigate conditions in which their competence fails. The models describing competence can then be used to pinpoint potential break points that cause people to make errors.

REFERENCES

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.

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

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cobos, P. L., López, F. J., Cano, A., Alvarez, J., & Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, **28**, 331-346.

De Houwer, J. (2002). Forward blocking depends on retrospective inferences about the presence of the blocked cue during the elemental phase. *Memory & Cognition*, **30**, 24-33.

De Houwer, J., & Beckers, T. (2002). Higher-order retrospective revaluation in human causal learning. *Quarterly Journal of Experimental Psychology*, **55B**, 137-151.

De Houwer, J., & Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *Quarterly Journal of Experimental Psychology*, **56B**, 345-357.

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, **55A**, 965-985.

Esmoris-Arranz, F. J., Miller, R. R., & Matute, H. (1997). Blocking of antecedent and subsequent events: Implications for cue competition in causality judgment. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 145-156.

Fenker, D., Waldmann, M. R., & Holyoak, K. J. (in press). Accessing causal relations in semantic memory. *Memory & Cognition*.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Science*, **7**, 43-48.

Goedert, K. M., & Spellman, B. A. (2005). Nonnormative discounting: There is more to cue interaction effects than controlling for alternative causes. *Learning & Behavior*, **33**, 197-210.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, **111**, 3-32.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, **30**, 1128-1137.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 97-106.

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

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, **395**, 71-73.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

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

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

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

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

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, **27**, 453-489.

Tangen, J. M., & Allan, L. G. (2004). Cue interaction and judgments of causality: Contributions of causal and associative processes. *Memory & Cognition*, **32**, 107-124.

Tangen, J. M., Allan, L. G., & Sadeghi, H. (2005). Assessing (in)sensitivity to causal asymmetry: A matter of degree. In A. Wills (Ed.), *New directions in human associative learning* (pp. 65-93). Hillsdale, NJ: Erlbaum.

Van Hamme, L. J., Kao, S. F., & Wasserman, E. A. (1993). Judging interevent relations: From cause to effect and from effect to cause. *Memory & Cognition*, **21**, 802-808.

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: Academic Press.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 53-76.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychological Bulletin & Review*, **8**, 600-608.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, **82**, 27-58.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 216-227.

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 López (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.

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

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, **10**, 92-97.

NOTES

1. The term *blocking* is not consistently used in the literature. For example, De Houwer et al. (2002) contrast blocking (i.e., certainty that the redundant cue is not a cause) with overshadowing (i.e., uncertainty about the causal status of the redundant cue). According to this terminology, uncertainty signifies absence of blocking (see also Williams, Sagness, & McPhee, 1994). However, this use collapses cases in which the redundant cue is rated lower than the predictive cue (but equal to the overshadowing control) with cases in which no difference is observed. We did not adopt this use because we wanted to differentiate full blocking (certainty that the redundant cue is not a cause or effect), partial blocking (uncertainty), and absence of blocking (certainty that the redundant cue is a cause or effect), which are all possible outcomes predicted by causal-model theory.

2. Our statistical conclusions are based on a significance level of .05 in all experiments. This value is also used for the interpretation of the theoretically derived specific comparisons, thus controlling the per-comparison error rate. However, we generally report the mean square errors and the descriptive p values (to three decimal places) (see also Goedert & Spellman, 2005, note 4). All our planned comparisons are based on the restricted error term. Thus, following the advice of Keppel and Wickens (2004, p. 520), we focus our analyses on the subset of the data we are currently analyzing. This strategy has the advantage of being more sensitive to possible differences of variance, and is in most cases also more conservative (because of the loss of degrees of freedom).

3. An interesting additional finding of this experiment is the fact that we found a blocking effect that differed from overshadowing (see note 1). The redundant cue was rated lower than the informative cues. Waldmann (2000, Experiment 1) found a similar effect in a predictive-learning condition using a similar task. According to a normative analysis (see introduction), participants should be as uncertain about the status of the redundant cue as they are uncertain about the two informative cues because the predictive cue is deterministic (see also De Houwer et al., 2002). Since this cue already generates an effect at the ceiling, the redundant cue cannot possibly display its power. We believe that the difference in the ratings is due to the fact that the redundant cue and the informative cues were both rated by the same participants, who intended to express that the redundant cue and the informative cues differed. For the redundant cue, which is paired with a deterministic cause, the data are consistent with the whole range of possibilities from being non-causal to also being a redundant deterministic cause. The most parsimonious assumption, which avoids redundant overdetermination, is to attribute the causal effect to the predictive cue only. In contrast, there is no reason to favor one of the two informative cues. Here the most parsimonious strategy is to divide causal strength between them or express equal uncertainty.