Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 2, 216–227

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.2.216

# Seeing Versus Doing: Two Modes of Accessing Causal Knowledge

Michael R. Waldmann and York Hagmayer
University of Göttingen

The ability to derive predictions for the outcomes of potential actions from observational data is one of the hallmarks of true causal reasoning. We present four learning experiments with deterministic and probabilistic data showing that people indeed make different predictions from causal models, whose parameters were learned in a purely observational learning phase, depending on whether learners believe that an event within the model has been merely observed ("seeing") or was actively manipulated ("doing"). The predictions reflect sensitivity both to the structure of the causal models and to the size of their parameters. This competency is remarkable because the predictions for potential interventions were very different from the patterns that had actually been observed. Whereas associative and probabilistic theories fail, recent developments of causal Bayes net theories provide tools for modeling this competency.

Causal knowledge underlies our ability to predict future events, to explain the occurrence of present events, and to achieve goals by means of actions. Thus, causal knowledge belongs to one of our most central cognitive competencies. However, the nature of causal knowledge has been debated. A number of philosophers and statisticians, such as Bertrand Russell (1913) and Karl Pearson (1892), have dismissed the notion of causality altogether and tried to replace it with the idea of correlation. This idea may be traced back to Hume (1748/1977), who argued that causality is an illusion based on associations that are produced by the experience of constant conjunctions of events. A modern variant of this approach is represented by theories that attempt to reduce causal learning to the acquisition of associative links between event representations (e.g., Shanks & Dickinson, 1987).

One fundamental problem of this view is that it collapses observational knowledge with interventional knowledge. Causal knowledge serves two different functions: It allows us to predict events on the basis of observed cues and at the same time underlies our ability to manipulate and control. For example, we can probabilistically predict the weather from readings of the barometer, and this prediction is driven by causal relations underlying the (spurious) covariation. Nevertheless, we also know that artificially setting the barometer to a specific reading would do nothing to the weather. Causal knowledge not only allows us to predict events on the basis of observed cues, it also tells us whether and which effects our actions will have. Although both types of prediction are driven by a common underlying causal model, the predicted outcomes may differ depending on whether the events are merely observed or actively set.

## Causal Versus Spurious Relations

Of course, psychological theories of causal induction did not completely disregard the fact that causal relations have to be distinguished from spurious relations. According to associative theories, such as the Rescorla-Wagner theory (Rescorla & Wagner, 1972), we typically learn associations in which the impact of potentially confounding cues is held constant. In some circumstances, this strategy may detect a spurious relation but it is far from fail-safe. For example, imagine an event A that is the cause of event B, which in turn is the cause of effect E. An associative account that handles both A and B as cues of outcome E would attempt to partial out the influence of A or B or both and therefore would completely miss the true causal model, which in this case is a causal chain. The reason for this failure is that these theories do not have the expressive power to represent causal models that differentiate between causes and effects and to adequately represent the structural implications of causal directionality (see also Waldmann, 1996). Even when causal models are provided to participants in the initial instructions, as in our experiments, the associative theories proposed in the literature are incapable of using this knowledge. Accordingly, a number of learning theorists endorsing the associative view have claimed that learners are insensitive to causal structure in trial-by-trial learning contexts regardless of whether participants are instructed about causal models or not (e.g., Cobos, López, Cano, Almaraz, & Shanks, 2002).

A second strategy of associationism is to differentiate between predictive and interventional knowledge altogether. Whereas classical conditioning might be viewed as underlying prediction, intervention might be driven by instrumental conditioning (Dickinson, 2001; see Domjan, 2003, for an overview). According to this view we might learn that the barometer reading predicts the weather (classical conditioning), and in a different setting we might additionally learn that interventions in the barometer are uncorrelated with the weather (instrumental learning). In this way we form separate associative weights for observational and interventional relations. However, although this approach approximates causal knowledge in many contexts, it fails to capture the relations between observations and interventions. The separation between

Michael R. Waldmann and York Hagmayer, Department of Psychology, University of Göttingen, Göttingen, Germany.

We thank A. Blaisdell, P. Cheng, C. Glymour, D. Lagnado, S. Sloman, and B. Spellman for helpful comments on a previous draft.

Correspondence concerning this article should be sent to Michael R. Waldmann or York Hagmayer, Department of Psychology, University of Göttingen, Gosslerstraβe. 14, 37073 Göttingen, Germany. E-mail: michael.waldmann@bio.uni-goettingen.de or york.hagmayer@bio.uni-goettingen.de

classical and instrumental conditioning predicts that without a prior instrumental learning phase, we are incapable of correctly predicting what would happen in case of an intervention in situations in which our knowledge is based on passive observations. Our experiments show that this is wrong.

One possible way to circumvent this problem is to postulate that associations based on classical conditioning can strengthen or weaken associations based on instrumental conditioning (see Rescorla & Solomon, 1967). Because in our experiments there is no instrumental (i.e., intervention) learning phase, this account is not viable. However, a simple extension would be to postulate that associative weights learned in the context of classical conditioning (i.e., observations) are transferred as initial weights to the instrumental learning system. We show below that this is also an inadequate account. Depending on the causal structure, predictions for observational cues and interventions should be similar or dissimilar to each other. We also show that learners prove capable of making correct inferences, which in some contexts do and in other situations do not mimic the associations learned in the observation phase.

Another approach to detect spurious relations is to control for co-factors when assessing causal strength (see Cheng & Novick, 1992; Spellman, 1996). This approach uses measures of covariation within subsets of events in which potential confounds are being held constant. One limitation of this approach is that it only applies to a specific class of causal models in which the target cause and the potential confounds are alternative, potential causes of a common effect (common-effect models; see Waldmann & Hagmayer, 2001).

Causal-model theory, which can be viewed as a psychological variant of Bayes net theories (Pearl, 1988; Waldmann & Martignon, 1998), is a more complete approach to dealing with these complexities (Waldmann, 2000, 2001; Waldmann & Holyoak, 1992; see also Rehder, 2003a, 2003b, for a similar view). According to this view, people acquire knowledge about complex causal models (e.g., common-effect models, chains, common-cause models) with directed causal links. The asymmetry of these links expresses our knowledge that causes precede and generate effects but not vice versa (see Waldmann, 1996). Causal-model theory postulates top-down learning as the main approach to the acquisition of causal knowledge. In most real-world situations people have prior assumptions about the causal status of events, which entails the direction of the causal arrow within causal models (see Waldmann, 1996). For example, we assume that switches are probably causes of lights and not vice versa even when we do not have more specific knowledge about how switches and lights are interrelated. Waldmann (1996) has discussed a number of sources for our prior assumptions (temporal cues, manipulations and interventions, verbal instructions, coherence with prior knowledge, etc.). Hypothetical causal models provide guidance for estimating the relevant parameters (e.g., causal strength, base rates). In this sense, causal-model theory is primarily a parameter estimation learning theory. However, Waldmann and Martignon (1998) have postulated that initial hypothetical models might be revised and modified when the mismatch between the causal model and the learning data becomes blatant.

Thus far, causal-model theory has only been tested in the context of observational predictive relations. It has been shown that causal models guide the strategies of estimating causal strength (Waldmann & Hagmayer, 2001) and that learners' predictions of events are sensitive to the causal structure connecting these events (Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne, 1995; Waldmann, 1996, 2000, 2001). Thus, people are sensitive to the difference between predictive cause–effect and diagnostic effect–cause relations and can access both relations for making predictions in an appropriate fashion. This competency proves that people are sensitive to the asymmetry between causes and effects when making predictions with observed events. However, the difference between predictions based on observations versus interventions has not been addressed within this approach to date. The difference between these two types of predictions can be seen best in diagnostic relations. Whereas it is possible to use observations of causal effects to reason back to their causes (diagnostic inference), it is not possible to generate the causes by manipulating their effects. Both inferences need to be sensitive to causal directionality but in different ways. The goal of the present research is to investigate whether people correctly differentiate between these two ways of accessing causal knowledge.

## Seeing Versus Doing: Theoretical Advances

One of the most important recent developments in the area of causality research are causal Bayes net theories (see Spirtes, Glymour, & Scheines, 1993; Pearl, 2000). Originally these theories have been developed as normative formal accounts of causal induction and causal inference, which could be implemented in machine learning algorithms. More recently, causal Bayes nets have also been proposed as the basis for psychological theories of human causal learning and reasoning (see Glymour, 2001, 2003; Gopnik et al., 2004; Lagnado & Sloman, 2004; Sloman & Lagnado, 2005; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003).

Causal Bayes net theories have many components, including algorithms for the induction of causal structures from covariation data. The aspect of these theories that is most relevant for the present studies involves the formal distinction between observing and intervening or seeing and doing (Pearl, 2000; Spirtes et al., 1993; Woodward, 2003). The key insight of this approach refers to the fact that the probability of an event conditional on the observation of other related events is not always identical to the probability of this event conditional on the intervention in these related events. Associative and probabilistic theories (including earlier Bayes net theories) did not have the conceptual power to distinguish between states of events that are observed as opposed to the very same states of events that are caused by interventions. The barometer example has already demonstrated that it is important to distinguish between these two types of states.

How can interventions be formally modeled? The most important component can be traced back to Fisher's (1951) analysis of experimental methods. Randomly assigning participants to experimental and control groups creates an independence between the independent variable and possible confounds. Thus, if the barometer is tampered with by free will, then the state of the barometer is independent of the pressure that typically affects it. To qualify as an intervention with this characteristic, the manipulation must force a value on the intervened variable (e.g., barometer), thus removing all other causal influences (e.g., atmospheric pressure). Moreover the intervention should be statistically independent of any variable that directly or indirectly causes the predicted event (e.g., all causes of weather), and it should not directly or indirectly cause the predicted event in addition to causing the intervened variable (see Pearl, 2000; Spirtes et al., 1993; Woodward, 2003, for formal analyses of interventions).

The insight that interventions with these proper characteristics create independence between intervened variables and their causes (i.e., possible confounds) underlies experimental methodology and is also one of the main components of current causal Bayes net theories. However, whereas experiments need to be run to test hypotheses, the new methods allow one to derive predictions about the effects of interventions even when the causal model underlying the predictions was induced on the basis of observational, nonexperimental data. Also whereas experimental methods typically deal with common-effect models with multiple causes leading to a common effect, the new Bayesian methods permit one to predict the outcomes of interventions in arbitrarily complex causal models. Spirtes et al. (1993) introduced the basic mathematical theory for predicting effects of "ideal" interventions in known causal systems, including various cases in which the causal structure is incompletely known. Pearl (2000) developed this work into a special "do" calculus that permits it to determine whether and how an effect of an intervention can be predicted.

To demonstrate the spirit of this analysis, Figure 1 shows two causal models that we have used in Experiment 1. Imagine the nodes in these networks representing substances in animals' blood whose level can either be increased or normal. In the top layer of Figure 1, P is a common cause of H and X; H, as well as X, causes G; and X also causes S. Imagine observational data has shown that all these relations (depicted by the causal arrows) are deterministic (i.e., increased levels cause increased levels) and that P is at an increased value with a base rate of 50%. The task is to predict the probability of an increased level of S conditional on H being present.

What are the predictions of causal Bayes nets when H is merely observed to be present? The left-hand side of Figure 1 shows the full causal model that is supported by the observations. If we observe that H is increased (i.e., in a "seeing" condition), then this allows us to infer back that P is also increased, which in turn allows us to infer that X, and therefore S, are also increased. Additionally, we can infer that G is increased. If H is on a normal level, then all other substances will also have normal values. This kind of reasoning is the domain of associative and probabilistic reasoning and can be modeled by networks that encode causal directionality or that encode all covariations between events.

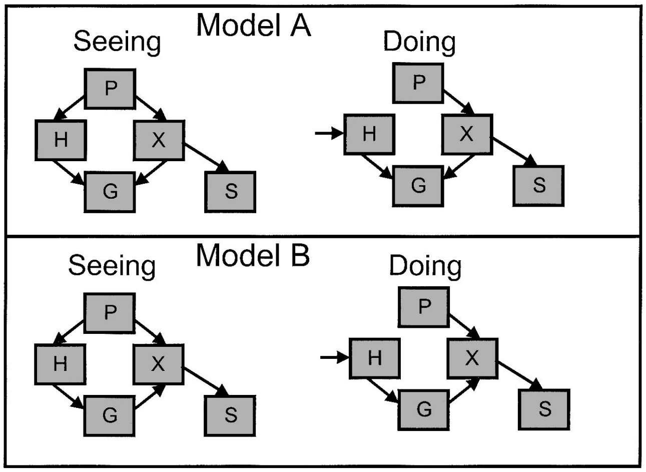

*Figure 1.* Two causal models (left side; A and B) presented in Experiment 1 along with observational data supporting deterministic relations. Intervening in event H normatively leads to the reduced models on the right side.

However, now imagine that the level of H is not observed but actively set by means of an intervention (e.g., an inoculation; i.e., a "doing" condition), and again the task is to predict the probability of an increased level of S. The right-hand side of Figure 1 illustrates this case. Because the intervention is determined by the free will of the agent (see the new arrow to H), the causal arrow between P and H must be deleted. According to Pearl's (2000) terminology, interventions entail "graph surgery." The intervention creates independence between the formerly causally connected events P and H. Although no data are available about actual interventions (i.e., no instrumental learning), it is still possible to make predictions about the outcome of hypothetical interventions. If H is manipulated, X, and in turn S, will be solely influenced by P, which happens to be increased with a base rate of 50%. Thus, events X and S will also be increased with a probability of .50. The probability of G will be codetermined by the now independent causes H and X. What is essential is that the probability of S conditional on an observed value of H does not equal the probability of S conditional on an intervention that forces that same value on H.

It is important to note that this analysis dissociates between observational and interventional predictions on the basis of identical observational data. It allows us to make predictions about patterns of events that may never have actually been observed (as in the intervention case in Figure 1). This feature goes beyond the capacity of probabilistic and associative theories and may be the single most important feature that makes causal Bayes nets truly causal.

Figure 1 (bottom row) shows a second condition we implemented in which the arrow between G and X is reversed. This condition serves as a control incorporating a causal structure, a causal chain, which does not entail different predictions for interventions and observations when the questions refer to increasing levels of H. This way it can be shown that learners do not always generate divergent predictions for interventions and observations. Because there is now a causal chain connecting H, G, X, and S and the causal relations are deterministic, increasing the level of H by means of an intervention will not lead to different predictions compared with merely observing an increased level of H. The level of S should be increased in both cases. However, different predictions result if H is observed to be at a normal level compared with the case in which H is forced to be at a normal level by an external intervention. Conditional on an observation of H being at a normal level, the probability of S having an increased level is zero in the data. However, if the level of H is actively set to a normal level, the probability of P having an increased level is still 0.5. Because P is an indirect deterministic cause of S, the probability of S being at an elevated level is 0.5 in the intervention condition.

## Seeing Versus Doing: Psychological Evidence

Given the recency of the conceptual advances within the literature on causal Bayes nets, only very few researchers have started to investigate the psychological validity of these models. A handful of recent studies have compared interventional with observational learning. The main goal of these studies was to investigate whether people can use covariational data to induce causal structure (i.e., structure learning) in a similar fashion as the machine algorithms that have been proposed in the literature (Pearl, 1988, 2000; Spirtes et al., 1993). No temporal or other cues were given so that the induction had to be based on the structure of probabilistic relations. For example, Steyvers et al. (2003) presented learners with data consistent with probabilistic common-cause, common-effect, and causal-chain structures and investigated the conditions under which the model could be correctly induced. The results demonstrated that learners showed some evidence of adequate learning when they merely observed the events but fared somewhat better, although far from ideal, when they were allowed to intervene (see Lagnado & Sloman, 2004, for similar results). With a similar goal Gopnik et al. (2004) have studied preschool children's ability to induce deterministic common-effect and common-cause structures on the basis of observing the effects of interventions. The results suggest that young children are also aided by the additional information interventions provide (see also Gopnik, Sobel, Schulz, & Glymour, 2001, for evidence that children choose causes but not spurious events for interventions). Thus, the main goal of these studies was to investigate the role of observations and interventions in inferring causal structure rather than the dissociation between predictions based on observations versus interventions.

In contrast to these studies, Sloman and Lagnado (2005) were interested in dissociating different types of inferences. They applied the causal Bayes net formalism to logical and counterfactual reasoning tasks. In a number of experimental studies they focused on various causal models (single causal relations, causal chains, and diamond-shaped causal structures) and interventions that removed events ("undoing"). The main focus of these studies was the comparison between inferences in causal scenarios and isomorphic conditional ("if–then") descriptions of these domains. The study most relevant for the topic of the present article is Experiment 2, which directly involved a case in which interventions and observations were dissociated. In this experiment, premises describing a probabilistic causal chain were used (e.g., "When A happens, it causes B most of the time. When B happens, it causes C most of the time. A happened. C happened."). The test questions in this study focused on the hypothetical absence of event B. The results showed that the inferences differed depending on whether participants were told that it was observed that B did not happen, or that B was actively prevented from occurring by means of an intervention. Participants were more likely to infer that A also happened in the intervention condition than to infer that it happened in the observation condition. Thus, there was a dissociation between responses to observation and intervention questions. The conclusions participants drew in this causal chain scenario were better accounted for by causal Bayes net theories than by traditional psychological theories of deductive and counterfactual reasoning.

## Goals of the Experiments

Our goal was to investigate whether people differentiate between intervening and observing in a task that involves estimating the parameters of given causal models. Moreover, relative to previous work in this area we increased the range of investigated causal models and the type of test questions. Unlike previous studies on learning, we were not interested in the question of whether people can induce the structure of causal models on the basis of covariations alone (i.e., structure learning; see Gopnik et al., 2004; Lagnado & Sloman, 2004; Steyvers et al., 2003). Following the framework of our top-down variant of causal-model theory, we assume that people typically acquire causal knowledge with at least minimal prior knowledge about the structure of the causal model (see Waldmann, 1996). This knowledge does not have to be specific; often we only use simple cues, such as temporal order (see also Lagnado & Sloman, 2004) or prior experiences with similar situations to decide whether events are potential causes or potential effects. Learning data is used to estimate the parameters of the causal model (i.e., causal strength, base rates of exogenous causes). In this regard, our research differs from Sloman and Lagnado's (2005) focus on reasoning with verbally described causal models. In our experiments participants receive instructions about the structure of different hypothetical causal models whose parameters are acquired in an initial learning phase. Most saliently in Experiments 3 and 4 we show that people's predictions are not only affected by the structure of the causal models but also by the parameters that were acquired in the learning phase. Moreover, we used a wider range of causal models than did Sloman and Lagnado (2005), and tested participants' inferences with test questions that did not only focus on the hypothetical absence of events (i.e., undoing) but also on their presence.

In summary, our main goal was to investigate whether people who went through a purely observational learning phase would be able to differentiate between predictions that are based on hypothetical observations (seeing) versus hypothetical interventions (doing). We were interested in finding out whether people can use identically acquired observational knowledge for deriving predictions and planning actions. A demonstration of this capacity would transcend the conceptual power of associative and probabilistic theories, which are restricted either to observational or to instrumental relations but cannot adequately capture both at the same time.

## Experiment 1

In this experiment we suggested the hypothetical causal models outlined in Figure 1 (left side) to learners and presented them with a list of individual learning cases that exhibited deterministic causal relations. After the learning phase, we asked participants to predict the probability of event S based on the hypothetical assumption that event H was either merely observed to be increased (seeing) or that its increased state was determined by an intervention (doing). As specified in the introduction, a normatively correct distinction between seeing and doing would entail differential answers for Model A (top row), whereas both questions should yield similar responses for Model B (bottom row) for the questions that refer to increased levels of H. For Model B both an intervention increasing the level of H and an observation of elevated levels of H imply an increased level of S. In contrast, for Model A only an observation of an increased level of H implies an elevated level of S but not an intervention on H, which should have no causal impact on S. In this condition, S is solely determined by its causes in the remaining model. Because P, the root cause in a causal chain, occurs with a base rate of 50% this should also be the prediction for S.

As a control condition, we also tested what participants would expect if H were observed to be at normal levels (seeing) or when the intervention decreased the level of H (doing). In contrast to the other condition, no dissociation between the models is expected in this scenario. The level of S should be expected to be normal in all test animals in the observation condition and at an elevated level in 50% of the animals in the intervention condition (Table 1, presented below, shows the normative numeric predictions in parentheses).

## Method

### Participants and Design

Fifty students from the University of Göttingen, Germany, participated in this experiment, with half of them randomly assigned to one of the two causal structures depicted in Figure 1 on the left side.

### Materials and Procedure

The cover stories, which were in German in all our experiments, presented a hypothetical causal model underlying sleeping sickness on the first page of a booklet. Participants were given either the model in the upper row or the model in the lower row. In the instruction for Model A, it was stated that scientists hypothesize that mosquito bites cause the production of the substance pixin. It is hypothesized that pixin causes substance xanthan, which causes the rise of the levels of both sonin and gastran. Pixin is also assumed to increase the level of histamine, which generates gastran. The instruction for Model B was almost identical, the only difference being that in this condition it was stated that the level of gastran increases the amount of xanthan and not vice versa. Participants were told that the researchers tested their hypothesis about the mechanism in a study using chimpanzees recently captured in Africa. In addition to the verbal descriptions of the causal relations, participants were shown a graph similar to the one in Figure 1. No abbreviations for the five substances were used in this graph.

On the second page participants received a list describing 20 chimpanzees that had been tested in a zoo. Each of the 20 cases was shown separately and conveyed information about the presence or absence of the five substances in the particular chimpanzee. Half of this group had increased levels of all five substances (P, H, X, G, S), half had normal values on all substances. Thus, all causal relations were deterministic, with P occurring with a base rate of 0.5. Participants were allowed to study the cases as long as they wished and to take notes. They were also permitted to refer back to the data and the graph of the causal model while answering the questions.

On the third page the test questions were listed. We gave participants two blocks of two questions each. One block consisted of the intervention questions (doing). The first of these questions asked participants to imagine that a doctor had inoculated 20 newly arrived chimpanzees with a substance that increased the level of histamine. The second intervention question stated that participants should assume that the substance decreased the level of histamine. The block with the observation questions (seeing) was similar, except for the fact that participants should imagine that 20 new chimpanzees were merely examined. The doctor had found that they all had either increased histamine levels or (in the other question) normal histamine levels. In all four questions participants were requested to estimate how many of the 20 chimpanzees will have an increased level of sonin. The sequence of the two blocks (seeing vs. doing), as well as the sequence of the two questions within the blocks, was counterbalanced.

### Results and Discussion

Table 1 shows the results of this experiment and the normative answers predicted by a causal Bayes net theory (in parentheses). An analysis of variance (ANOVA) with causal model (A vs. B) as a between-subjects variable and the hormone level (increased vs. normal) and type of question (seeing vs. doing) as within-subject variables revealed a significant three-way interaction, $F(1, 48) = 19.99$, $p < .01$, which was followed up with more focused tests. First, an ANOVA was conducted for the observation questions with causal model and level as factors. Consistent with our prediction there was no significant difference between the responses to the observation questions (seeing) across the two causal models (A and B), ($F < 1$). The correct answer would be the prediction of increased levels of sonin when histamine is at an increased level and normal levels when it is at a normal level; the participants' ratings, which could range between 0 and 20, clearly reflect this pattern, $F(1, 48) = 222.14$, $p < .01$. There was no interaction between the causal model and the level of H ($F < 1$). Participants' answers deviated only slightly from the normative predictions.

In contrast to the observation questions, a normative account predicts a dissociation between the two models for the intervention questions (doing), which was actually found. The results of an ANOVA for these questions, with causal model and level of H used as factors, showed the expected strong effects for the factors model, $F(1, 48) = 15.49, p < .01$, and level, $F(1, 48) = 73.51, p < .01$. The interaction also proved significant, $F(1, 48) = 25.60, p < .01$. Although the statistical patterns conform to the normative predictions, participants' answers to some of the intervention questions deviated from the normative predictions. The predictions for interventions that increased substance H were fairly close to the normative values. Here, it is particularly noteworthy that the predictions for the crucial model A correctly were placed in between the observational answers at a value that had never been observed in the data. This pattern convincingly demonstrates that participants were capable of dissociating intervening and observ-

Table 1
*Means and Standard Deviations of the Responses to the Two Observation Questions (Seeing) and the Two Intervention Questions (Doing) in Experiment 1*

| Causal model | Intervention | | Observation | |
|---|---|---|---|---|
| | Increasing | Lowering | Increased | Normal |
| Model A | | | | |
| *M* | 8.48 (10) | 4.84 (10) | 17.44 (20) | 2.96 (0) |
| *SD* | 6.56 | 4.99 | 4.48 | 6.08 |
| Model B | | | | |
| *M* | 17.72 (20) | 3.60 (10) | 17.40 (20) | 2.32 (0) |
| *SD* | 3.92 | 4.68 | 4.11 | 3.98 |

*Note.* Normative responses (range = 0–20) are presented in parentheses.

ing. The predictions for the intervention that decreased the substance were too low, however. A possible explanation might be that participants had trouble figuring out the consequences of a decrease of H in the different possible scenarios. They never had observed cases in which normal levels were decreased. Some participants actually mentioned that they thought that decreasing the level of H below a normal level might also reduce S to a subnormal level when it was expected to be normal prior to the intervention.

Overall the results demonstrate that participants proved capable of accessing causal models whose parameters were learned on the basis of observational data differently depending on whether they had to derive predictions for potential observations or for potential interventions. They did not simply access the observed associative or probabilistic relations but seemed to have correctly accessed the underlying causal model before making predictions in a task-specific fashion. The responses are not perfect but clearly reflect the competency to distinguish between seeing and doing.

A critical question concerning the results of this first experiment is whether participants in fact used the parameters they had learned to make their predictions. An alternative interpretation of the results might be that the participants only considered the causal model and assumed a default base rate of 0.5 but that they did not use the specific parameters inherent in the data. The next three experiments were designed to show that participants do learn the parameters and use them to derive predictions for interventions.

## Experiment 2

The first experiment presented participants with a hypothetical causal model and learning data that fleshed out the strength and the direction of covariation generated by the causal arrows. These data supported deterministic causal relations with positive correlations and a base rate of the root cause of the target effect of 0.5. Our goal in the second experiment was to test whether the dissociation between seeing and doing can be replicated with probabilistic causal models. In this experiment we contrasted a common-cause model with a causal-chain model (see Figure 2) and presented data that exhibited probabilistic relations. We used identical learning data for both models, implying identical probabilistic relations between the cause-and-effect events in both models. A normative account would imply no difference between predictions that are based on observations for the two models but clear differences between predictions of the outcomes of hypothetical interventions. Whereas seeing questions should yield different responses than doing questions in the common-cause model, similar responses are expected in the causal chain model.

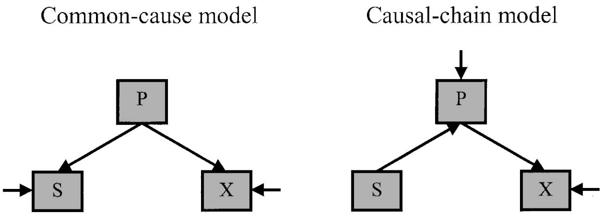

*Figure 2.* Common-cause and causal-chain models presented in Experiment 2 along with observational probabilistic data. The arrows pointing from outside the models represent unknown factors.

### Method

### Participants and Design

Forty-eight students from the University of Göttingen participated in this experiment. They were randomly assigned to the common-cause or the causal-chain condition.

### Materials and Procedure

We used the same cover story as in Experiment 1 but mentioned only the (fictitious) substances sonin (S), pixin (P), and xanthan (X). The instruction displayed either the common-cause or the causal-chain model depicted in Figure 2. Participants were either told that pixin is assumed to cause both sonin and xanthan or that sonin presumably causes pixin, which causes xanthan. To alert participants to the possibility of probabilistic relations, it was pointed out that other unknown causal factors might stimulate or inhibit the formation of the substances. We added additional arrows representing these alternative unobserved causes to the graphs in the instruction.

On the second page we presented the data. Participants received information about 20 individual chimpanzees. As before, the cases were described in a list of cases, which participants could study as long as they wished. The statistical structure of the learning data was identical in both conditions. Eight of the chimpanzees had elevated levels of all three substances, and eight had normal levels of all of them. Four animals had patterns of substances inconsistent with deterministic relations: One had elevated levels of P and X and a normal level of S, another had normal levels of P and X and an increased level of S, a third had increased levels of P and S and a normal level of X, and the fourth had normal levels of P and S and an increased level of X.

Given a common-cause model, these data imply that the two effects S and X are at an increased level with a probability of .9 and .1 conditional on the cause P being at an increased or normal level, respectively. The base rate of P was .5. The very same data imply in the causal-chain condition that the intermediate effect and the final effect are at an increased level with a probability of .9 when the respective direct causes (i.e., the initial cause or the intermediate event) were at increased levels, and .1 otherwise. The base rate of S was .5.

On the basis of these parameters and the causal model provided in the instructions, a normative Bayesian analysis can be conducted (see Appendix for details). This analysis implies for both models that the target effect X has a probability of .82 when S is observed to be at an increased level and a probability of .18 when it is observed to be at a normal level. These probabilities slightly deviate from the probabilities given in the data (.8 and .2). The differences are due to rounding errors, which cannot be avoided in a sample with 20 cases. For the causal-chain model the normatively implied observational probabilities are identical to the interventional probabilities. However, intervening in S in the common-cause model entails the removal of the arrow connecting P and S as a result of graph surgery. This means that the probability of hormone X being increased still depends on the base rate of P and the strength of the causal relation between P and X but is now independent of the level of S. On the basis of the given data, the probability of X being increased is 0.5 regardless of whether the intervention in S increases or decreases the level of S.

The test questions (seeing vs. doing) were presented on a third sheet. The questions and the counterbalancing were the same as in Experiment 1. We either asked participants to imagine 20 new chimpanzees whose level of hormone S had been increased or decreased by means of an intervention (doing) or who only happened to have increased or normal levels (seeing). The task of the participants was to estimate the number of animals who would have high levels of substance X in these four different conditions. The normative answers to the four questions are shown in Table 2 in parentheses. The expected frequencies are rounded to the next integer.

Table 2
*Means and Standard Deviations of the Responses to the Two Observation Questions (Seeing) and the Two Intervention Questions (Doing) in Experiment 2*

| | Intervention | | Observation | |
|---|---|---|---|---|
| Causal model | Increasing | Lowering | Increased | Normal |
| Common-cause model | | | | |
| *M* | 9.04 (10) | 6.29 (10) | 14.79 (16) | 3.29 (4) |
| *SD* | 5.86 | 4.89 | 4.10 | 1.81 |
| Causal-chain model | | | | |
| *M* | 14.04 (16) | 3.08 (4) | 13.67 (16) | 3.08 (4) |
| *SD* | 5.07 | 3.71 | 4.89 | 1.89 |

*Note.* Normative responses (range = 0–20) are presented in parentheses.

### Results and Discussion

As can be seen in Table 2, the results closely mimic the ones in Experiment 1. An ANOVA with causal model (common-cause vs. causal-chain) as a between-subjects variable and hormone level (increased vs. normal) and type of question (seeing vs. doing) as within-subject variables revealed a significant three-way interaction, $F(1, 46) = 18.93$, $p < .01$. Again there were no significant differences in the responses to the observation questions (seeing) across the two causal models (common-cause vs. causal-chain). An ANOVA of the responses to the observation questions with the level of S and causal model variables yielded only a strong main effect of level, $F(1, 46) = 272.51$, $p < .01$ (all other $Fs < 1$). In fact, the responses to the observation questions in both conditions and to the intervention questions in the causal-chain condition, which normatively should be similar, were indeed very similar (see Table 2).

A normative Bayesian analysis predicts responses to the intervention questions in the common-cause condition that are very different from the responses in the causal-chain condition. The pattern of predicted normative responses (see Table 2) entails an interaction between the causal model and the level of S. It also predicts a main effect of level but not of causal model. An ANOVA of the responses to the intervention questions with causal model and level of S as variables clearly supported these predictions. There was a strong interaction, $F(1, 46) = 17.02$, $p < .01$, and a main effect for level of S, $F(1, 46) = 47.46$, $p < .01$. As expected, the factor causal model was not significant ($F < 1$).

Minor deviations from the normative predictions were again found when the hypothetical intervention decreased the level of X. Participants underestimated the probability of increased levels of X if the level of S was decreased by an intervention. We think that participants had similar problems as in Experiment 1 with figuring out the consequences of a decrease of normal levels.

Overall the results replicate Experiment 1 with different causal models and probabilistic data. The results show that people not only distinguish between seeing and doing but also that they indicate sensitivity to the probabilistic nature of the relations. Comparing the estimates for the deterministic relations in Experiment 1 with the estimates in Experiment 2 in the observation conditions shows that individuals apparently learned different causal strength estimates. Thus, participants seemed to have acquired different parameters, which influenced the predictions. This finding weakens the alternative interpretation of the results of Experiment 1 that learning might not have been involved in participants' prediction responses.

Although the responses were again not perfect (and descriptively deviated in the same direction as in Experiment 1), these deviations were smaller. It may have been easier to reason with causal models that only contain three events than with those that contain five events, as in Experiment 1.

## Experiment 3

The aim of the following two experiments was to further investigate the role of the learning data. In all our experiments, people started learning with hypothetical causal models and had to use learning data to fill in the parameters of the causal models (i.e., base rates of causes, causal strength of links). In Experiments 1 and 2 we used an intermediate base rate of 0.5 and either equal deterministic relations (Experiment 1) or equal probabilistic relations (Experiment 2) for the different causal links. Although the results of the experiments suggest sensitivity to both the instructed causal model and the parameters, the argument could still be made that the inferences might mainly be driven by the structure of the hypothetical causal models that were initially given to the learners and only to a minor extent by the learned parameters. The contrast between Experiments 1 and 2 already provided suggestive evidence that people attended to the learning data; the predictions exhibited sensitivity to the differences in the strength of causal relations across the two experiments. The main goal of Experiments 3 and 4 was to provide unambiguous evidence that participants integrated the parameter estimates in their predictions. To investigate sensitivity to the parameters, we manipulated base rates in Experiment 3 and causal strength in Experiment 4.

Base rates are especially important for predictions within a common-cause structure. If one of the effects (e.g., effect 1) within such a structure is manipulated by an intervention, the probability of the second effect depends solely on the base rate of the common cause and the strength of the causal connection between the cause and the second effect (e.g., effect 2). In contrast, if the presence of one of the effects is merely observed, the probability of the second effect still depends on the strength of this causal relation but also on the posterior probability of the cause given that the first effect is present. According to the Bayes rule, this posterior probability is dependent on the base rate of the cause and the strength of the causal relation between the common cause and the first effect. Therefore, different base rates of the common cause have a stronger influence on the probability of the second effect in the context of interventions in the first effect than in the context of observations of the first effect (see the Appendix for formal details). In intervention contexts, the probability of the second effect rises proportional to the base rate of the common cause, and there should be no difference between generating and preventing the first effect. In contrast, in observation contexts, the probability of the second effect will only rise slightly when the base rate of the common cause is increased, and there will remain a difference between the presence and absence of the first effect.

Experiment 3 was designed to test whether participants were aware of the differential impact of different base rates. Participants were given a common-cause structure in which the common cause either had a high or a low base rate (.8 vs. .2). The causal relations between the cause and both effects were probabilistic and identical in all conditions.

### Method

### Participants and Design

Thirty-two students from the University of Göttingen participated in this experiment. They were randomly assigned to either the high or the low base rate condition.

## Materials and Procedure

In both conditions, a common-cause structure was presented in the initial instructions. Participants were told on the first page of a booklet that researchers had discovered a new bacterium in dogs (common cause), which they assumed to be causing gastric problems (effect 1) and the presence of antibodies (effect 2). Participants were also informed that there are other possible causes of the two effects, acidic substances that might cause gastric problems and viral infections that might lead to the formation of antibodies. As in the previous experiments, the causal relations were visualized by a graph.

On the second page, participants received data from a fictitious study about gastric problems of dogs. Again the cases were presented individually on a list. Test questions were given on the next two pages in the booklet. One page was reserved for the observation questions and one for the intervention questions. The order of the test questions was counterbalanced as in the previous experiments.

In the test phase, participants were asked to imagine 20 dogs who had previously not been tested as either all having gastric problems or all not having gastric problems (seeing) and to imagine 20 dogs that were all given either a substance that causes gastric problems or a substance that cures these problems in all dogs (doing). Participants had to estimate the number of animals having antibodies in these four conditions.

The data set consisted of 20 cases. Each dog was tested for the bacterium, gastric problems, and antibodies. In the high base rate condition, 16 of the 20 dogs were infected, 14 of these animals also had gastric problems and antibodies, 1 had no gastric problems but antibodies, and 1 had antibodies but no gastric problems. The 4 uninfected dogs all had no stomach trouble and no antibodies. In contrast, in the low base rate condition only 4 dogs were infected, which had both gastric problems and antibodies. Of the 16 uninfected animals 14 had neither stomach problems nor antibodies, 1 had only gastric problems and 1 showed only antibodies. These data imply in both conditions a contingency of $\Delta P = .94$ between the common cause and each of the effects ($\Delta P$ is defined as the difference between the conditional probabilities $P$[effect|cause] and $P$[effect|noncause]). A normative Bayesian analysis based on the equations shown in the Appendix implies distinct probabilities of the second effect for the observation and intervention questions. Table 3 lists the normative predictions based on the model and the parameters derived from the data. Table 4 shows the expected frequencies rounded to the next integer.

## Results and Discussion

The results and the normative predictions for the frequency ratings are displayed in Table 4. We conducted an ANOVA with estimates of the number of animals having antibodies as the dependent variable, the base rate (high vs. low) as a between-subjects variable, and the type of question (seeing vs. doing) and presence of first effect (present vs. absent) as within-subject variables. This analysis yielded significant main effects for the presence of first effect and base rate variables, $F(1, 30) = 132.90$, $p < .01$, and $F(1, 30) = 8.83$, $p < .01$, respectively. These two effects indicate overall sensitivity to the base rate manipulation and to the differences between present and absent effect cues. The three-way interaction was not significant, but there was a significant interaction between the presence of first effect and type of question variables, $F(1, 30) = 46.18$, $p < .01$, indicating that participants again differentiated between seeing and doing. The interaction between the base rate and presence of first effect variables was also significant, $F(1, 30) = 5.55$, $p < .05$.

Table 3
*Normatively Implied Probabilities of the Second Effect (e2) Conditional on the First Effect (e1) Being Observed or Manipulated (Experiment 3)*

| Base rate | Intervention | | Observation | |
|---|---|---|---|---|
| | $P(e2\|do[e1])$ | $P(e2\|do[\sim e1])$ | $P(e2\|e1)$ | $P(e2\|\sim e1)$ |
| High base rate<br>$P_{\text{infection}} = .80$ | .75 | .75 | .94 | .19 |
| Low base rate<br>$P_{\text{infection}} = .4$ | .25 | .25 | .81 | .06 |

Table 4
*Means and Standard Deviations of the Responses to the Two Observation Questions (Seeing) and the Two Intervention Questions (Doing) in Experiment 3*

| Base rate | Intervention | | Observation | |
|---|---|---|---|---|
| | Effect present | Effect absent | Effect present | Effect absent |
| High base rate | | | | |
| *M* | 13.31 (15) | 9.00 (15) | 18.19 (19) | 2.06 (4) |
| *SD* | 5.75 | 6.48 | 1.68 | 1.53 |
| Low base rate | | | | |
| *M* | 8.31 (5) | 5.94 (5) | 13.50 (16) | 2.37 (1) |
| *SD* | 5.81 | 5.31 | 6.35 | 1.75 |

*Note.* Normative responses (range = 0–20) are presented in parentheses.

To explore the results in greater detail we analyzed the intervention and observation data separately. The analysis of the intervention data yielded a significant effect of base rate, $F(1, 30) = 5.68$, $p < .05$, which demonstrates that the predictions were sensitive to the base rates. As predicted, the interaction was not significant ($F < 1$), but there was a significant effect of the presence of first effect variable, $F(1, 30) =7.89$, $p < .01$, which deviates from the normative predictions. As in the previous experiments, there was a tendency to give lower ratings when the effect cue was absent. This finding is likely due to some participants giving observation responses to the intervention questions. Especially when the base rates are high, it seems difficult to ignore an intervention that removes an effect when predicting the second effect. An alternative explanation might be that some participants induced an additional causal link between the first and the second effect. However, this explanation is weakened by the fact that according to everyday knowledge, organisms do not generate antibodies against substances.

The corresponding analyses of the observation data also showed a significant effect of the base rates, $F(1, 30) = 6.79$, $p < .05$, which is consistent with the small difference that the normative analysis predicts. There was also a large effect of the presence of first effect variable, $F(1, 30) = 229.22$, $p < .01$, which corresponds to the large difference predicted by the normative analysis. Finally, the interaction unexpectedly proved significant, $F(1, 30) = 7.72$, $p < .05$, which, possibly because of regression effects, reveals a pattern that slightly deviates from the normative prediction. Despite these relatively minor deviations from normative responses, the general pattern clearly demonstrates that people are sensitive to base rates when making predictions in the contexts of seeing and doing.

# Experiment 4

In Experiment 4, we manipulated another important parameter of causal models, the strength of the causal relations. As in Experiment 3, we focused on a common-cause model. Two conditions were compared: In one condition (strong–weak), the causal connection between the common cause and the first effect was strong ($\Delta P = .91$), and the causal connection to the second effect was weak ($\Delta P = .45$). In the second condition (weak–strong), this assignment was reversed without changing the base rate of the common cause. This comparison allowed us to empirically test whether people use causal strength when making predictions in seeing and doing contexts. If there is an intervention in the first effect, effect 1, the probability of the second effect (effect 2) depends on the base rate of its cause and the strength of the causal relation between the cause and effect 2. Therefore the presence of the second effect is more likely in the weak–strong condition than in the strong–weak condition independent of the type of intervention in effect 1. If on the other hand, the first effect is merely observed, the probability of the second effect varies depending on the presence or absence of the first effect. In our task, both probabilities are higher in the weak–strong condition than in the strong–weak condition. On the basis of the previous results, we expected participants to be sensitive to the implications entailed by the parameters of the different conditions.

## *Method*

### *Participants and Design*

Thirty-two students from the University of Göttingen participated and were randomly assigned to one of two experimental conditions, which manipulated the strength of the causal relations between a common cause and its two effects. Either the relation to the first effect was strong and the relation to the second effect was weak (strong–weak condition) or vice versa (weak–strong condition).

### *Materials and Procedure*

The instructions and the procedure were adopted from Experiment 3. Participants were again given the task to study data about a hypothetical causal model relating a bacterial infection (cause) to gastric problems (effect 1) and the presence of antibodies (effect 2) in dogs. As in Experiment 3, a sample of 20 individual cases was shown to participants as learning data. The same test questions were used as in the previous experiment with the order again being counterbalanced. Participants were asked to imagine a sample of 20 new cases in which gastric problems were observed to be either present or absent or were generated or prevented by means of an intervention. The task was to predict the number of dogs having antibodies in these four different conditions.

The data were again described as showing the results of a study about gastric problems of dogs. Again, the cases were presented on a list, which showed whether the causal events were present or absent in the particular dog. The dogs were individually tested for the bacterium, gastric problems, and antibodies. In the strong–weak condition, 5 animals showed bacteria, gastric problems, and antibodies; another 5 dogs had bacteria and gastric problems but no antibodies; 1 dog had only a bacterial infection without any effects; and the remaining 9 dogs were uninfected and also showed none of the effects. In the weak–strong condition, 5 dogs were infected and had both gastric problems and antibodies, 5 dogs had bacteria and antibodies but no gastric problems, 1 dog was infected and showed no effects, and 9 dogs had neither bacteria nor any of the effects. The frequencies showed that both data sets were completely symmetric. The resulting contingencies were $\Delta P = .91$ for the strong and $\Delta P = .45$ for the weak causal relation. The base rate of the cause was $P_{\text{bacteria present}} = .55$ in both data sets. Table 5 lists the probabilities normatively implied by the parameters derived from the data set (see the Appendix for the formulas). The rounded normatively expected frequencies of the second effect in the presence or absence of the first effect in all eight conditions are shown in parentheses in Table 6.

### *Results and Discussion*

Table 6 shows the results for the frequency estimates. An ANOVA was conducted including causal strength (strong–weak vs. weak–strong) as a between-subjects variable and type of question (seeing vs. doing) and presence of first effect (present vs. absent) as within-subject variables. The analysis yielded significant main effects for the presence of first effect, $F(1, 30) = 31.87$, $p < .01$, and causal strength, $F(1, 30) = 22.21$, $p < .01$, variables. Both effects are predicted by the normative analysis. Table 5 shows that the normatively implied probabilities of the second effect are always lower in the strong–weak condition than the corresponding probabilities in the weak–strong condition. As in the previous experiment, there was a significant interaction between presence of first effect and type of question, $F(1, 30) = 41.40$, $p < .01$, which demonstrates once again that participants differentiated between observation and intervention. There was also an interaction between the type of question and the strength of the causal relations, $F(1, 30) = 8.23$, $p < .01$. Moreover, the three-way interaction turned out to be marginally significant, $F(1, 30) = 3.77$, $p < .10$.

To further explore these interactions, we conducted separate analyses for the intervention and observation data. The analyses of the intervention data showed that neither the presence of first effect variable nor the interaction was significant. This is consistent with the normative prediction, although the data indicated a tendency to show regression effects. The predicted difference of the causal strength variable was significant with a one-tailed test, which seems appropriate given our specific prediction, $F(1, 30) = 3.17$, $p < .10$ (two-tailed).

The statistical analyses of the observation data conformed to the normative predictions. There was a significant difference between the conditions in which the effect cue was described to be present or absent, $F(1, 30) = 67.36$, $p < .01$, as well as the predicted strong effect for the factor causal strength, $F(1, 30) =30.98$, $p < .01$. The interaction was not significant.

In summary, again we showed that people differentiated between seeing and doing. Whereas Experiment 3 demonstrated that participants were sensitive to the base rates of the common cause when making predictions in these two contexts, the present experiment extends this finding to the parameter of causal strength.

## General Discussion

The ability to derive predictions for the outcomes of potential actions from observational data is one of the hallmarks of true causal reasoning (Pearl, 2000; Spirtes et al., 1993). The current four experiments showed that people indeed have the competency to derive different predictions from an observationally acquired causal model depending on whether they believed that a causal

Table 5
*Normatively Implied Probabilities of the Second Effect (e2) Conditional on the First Effect (e1) Being Observed or Manipulated (Experiment 4)*

| | Intervention | | Observation | |
|---|---|---|---|---|
| Causal relations | $P(\text{e2}\mid\text{do}[\text{e1}])$ | $P(\text{e2}\mid\text{do}[\sim\text{e1}])$ | $P(\text{e2}\mid\text{e1})$ | $P(\text{e2}\mid\sim\text{e1})$ |
| Strong–weak | | | | |
| $\Delta P_{\text{c-e1}} = 0.91$<br>$\Delta P_{\text{c-e2}} = 0.45$ | .25 | .25 | .45 | .05 |
| Weak–strong | | | | |
| $\Delta P_{\text{c-e1}} = 0.45$<br>$\Delta P_{\text{c-e2}} = 0.91$ | .50 | .50 | .91 | .36 |

Table 6
*Means and Standard Deviations of the Responses to the Two Observation Questions (Seeing) and the Two Intervention Questions (Doing) in Experiment 4*

| Causal relations | Intervention | | Observation | |
|---|---|---|---|---|
| | Effect present | Effect absent | Effect present | Effect absent |
| Strong–weak | | | | |
| *M* | 7.31 (5) | 5.88 (5) | 7.81 (9) | 1.31 (1) |
| *SD* | 3.52 | 5.14 | 4.00 | 2.87 |
| Weak–strong | | | | |
| *M* | 8.56 (10) | 8.60 (10) | 15.3 (18) | 5.85 (7) |
| *SD* | 4.73 | 2.57 | 6.07 | 2.44 |

*Note.* Normative responses (range = 0–20) are presented within parentheses.

event had been merely observed or was actively manipulated. Whereas associative learning theories often have tried to circumvent the problem of distinguishing between seeing and doing by postulating two separate learning mechanisms for these two types of events (classical and instrumental conditioning) or by assuming simple transfer, the present results show that people can derive instrumental predictions without actually having undergone an instrumental learning phase. This competency is remarkable because the predictions derived for potential interventions were often very different from the patterns that were actually observed. For example in Experiment 1, learners observed deterministic relations that they could easily use to answer the observation questions. But in the appropriate causal model, they also were able to predict an intermediate probability of the target event when they had been asked to assess the outcomes of hypothetical interventions. This capability was evident despite the fact that the probability of the target event was never actually observed. We also used a causal chain model that does not imply a dissociation between seeing and doing when the root of the chain is used as the predictor and found that participants indeed gave similar ratings for both types of questions. This shows that learners do not generally lower their ratings when predicting the outcomes of interventions but are sensitive to the normative differences between causal models. Normative accounts of causality (i.e., causal Bayes nets) can model the distinction between seeing and doing by postulating a learning phase of the complete causal model and (if necessary) subsequent graph surgery before predictions for interventions are derived (Pearl, 2000; Spirtes et al., 1993).

Unlike objectives of previous research, we were not interested in investigating how people acquire causal models from scratch (e.g., Gopnik et al., 2004; Lagnado & Sloman, 2004; Steyvers et al., 2003). Following causal-model theory (see Hagmayer & Waldmann, 2002; Waldmann, 1996), we believe that in virtually all realistic learning situations, people bring to bear prior knowledge about the structure of causal models. This knowledge may be rudimentary and hypothetical. Often a distinction between potential causes and effects based on cues (e.g., temporal order, instructions, analogies, interventions) suffices. Thus, our account of learning assumes that people have initial assumptions about hypothetical causal models and that they fill in the details about the parameters when presented with data. This is a form of top-down learning, which does not simply involve reasoning with instructed causal models (as in Sloman & Lagnado, 2005), because it can be shown that knowledge about parameters is needed and used to make correct predictions. Experiments 3 and 4, in particular, have demonstrated that participants were sensitive to differences of the parameters estimated on the basis of observed learning input. It could be shown that base rate and causal strength parameters were encoded and used in the predictions.

Of course, the normative theories we were exploiting for our predictions (Pearl, 2000; Spirtes et al., 1993) are capable of far more complex inferences than the ones we requested from our participants. These theories, which were originally developed for machine learning, allow one to derive predictions for very complex models or to determine which additional variables must be measured in order to make predictions possible. In addition, techniques have been developed that allow machines to induce causal models from merely covariational data or mixtures of observations and interventions. Future research will be needed to determine the number of features shared by people's intuitive reasoning about causal models. A number of recent studies suggest that there are boundary conditions for people's competencies to learn and reason according to these normative models (Lagnado & Sloman, 2004; Waldmann & Walker, in press).

Some of our results showed that people are not always perfect when making predictions. In the intervention conditions, they had a tendency to underestimate the causal role of the causes that were still influencing the target effect with a specific base rate. Moreover, the data in the first three experiments exhibited somewhat greater difficulties with interventions that prevented events than ones that generated them. Some of the participants also may have confused interventions with observations or tended toward regressive responses when their confidence was low. In general, predictions about interventions seemed harder than predictions based on observation. This is not surprising because the former is computationally more complex than the latter. In both tasks, covariations between events must be encoded. Whereas the observation task allows for direct use of these observed probabilities, the intervention questions could only be correctly answered in some conditions if the underlying causal model was altered and patterns were inferred from the modified causal models that were in conflict with the observations.

The factors causing deviations from normative responses need to be explored further in future research. Moreover, it would be interesting to investigate whether other learning tasks (e.g., trial-by-trial learning), other domains, and other populations of participants show similar results. Nevertheless, the present results demonstrate people's remarkable competency to engage in true causal reasoning and learning.

## References

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

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

Dickinson, A. (2001). Causal learning: An associative analysis. *Quarterly Journal of Experimental Psychology, 54B,* 3–25.

Domjan, M. (2003). *The principles of learning and behavior* (5th. ed.). Belmont, MA: Thomson/Wadsworth.

Fisher, R. (1951). *The design of experiments*. Edinburgh: Oliver & Boyd.

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.

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.

Gopnik, A., Sobel, D. M., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37,* 620–629.

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

Hume, D. (1748/1977). *An enquiry concerning human understanding.* Indianapolis, IN: Hackett.

Lagnado, D., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 856–876.

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

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

Pearson, K. (1892) *The grammar of science*. London, England: Walter Scott.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27,* 709–748.

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1141–1159.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian and instrumental learning. *Psychological Review, 74,* 151–182.

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.

Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society, 13,* 1–26.

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

Sloman, S. A., & Lagnado, D. (2005). Do we "do"? *Cognitive Science, 29,* 5–39.

Spellman, B. A. (1996). Conditionalizing causality. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 167–206). San Diego, CA: Academic Press.

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.

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

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

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychological Bulletin and 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., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121,* 222–236.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124,* 181–206.

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

Waldmann, M. R., & Walker, J. M. (in press). Competence and performance in causal learning. *Learning and Behavior.*

Woodward, J. (2003). *Making things happen. A theory of causal explanation.* Oxford, England: Oxford University Press.

# Appendix

## Modeling Observations and Interventions in Causal Bayes Nets

The following analyses apply to common-cause models or causal chains with three events (see Figure 2). Capital letters represent random variables of events, small letters represent instantiations of events. For example, el indicates that event E1 is present and ~e1 indicates that E1 is absent.

### Common-Cause Model

A common-cause model with three events describes a situation in which a common cause C influences two effects E1 and E2. Because of the Markov condition, which screens off the two effects when their common cause is held constant, the joint probability can be factored with the use of the expression

$$P(\mathrm{E2,E1,C}) = \mathrm{P}(\mathrm{E1}|\mathrm{C}) \times \mathrm{P}(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}). \qquad (1)$$

#### *Observation*

Given that E1 is observed to be present (E1 = e1), the model implies that

$$P(\mathrm{E2}|\mathrm{e1}) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|\mathrm{e1}). \qquad (2)$$

If E1 is observed to be absent, the resulting equation is

$$P(\mathrm{E2}|{\sim}\mathrm{e1}) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|{\sim}\mathrm{e1}). \qquad (2')$$

Unpacking equations 2 and 2′ yields

$$P(\mathrm{e2}|\mathrm{e1}) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}|\mathrm{e1}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P( \sim \mathrm{c}|\mathrm{e1})$$

$$P(\mathrm{e2}|{\sim}\mathrm{e1}) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}|{\sim}\mathrm{e1}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c}|{\sim}\mathrm{e1}).$$

According to the Bayes rule these equations are equivalent to

$$P(\mathrm{e2}|\mathrm{e1}) = [P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{e1}|\mathrm{c}) \times P(\mathrm{c}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P(\mathrm{e1}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c})]/[P(\mathrm{e1}|\mathrm{c}) \times P(\mathrm{c}) + P(\mathrm{e1}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c})].$$

$$P(\mathrm{e2}|{\sim}\mathrm{e1}) = [P(\mathrm{e2}|\mathrm{c}) \times P({\sim}\mathrm{e1}|\mathrm{c}) \times P(\mathrm{c}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{e1}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c})]/[P({\sim}\mathrm{e1}|\mathrm{c}) \times P(\mathrm{c}) + P({\sim}\mathrm{e1}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c})].$$

These equations allow one to calculate the probability of the presence of the second effect given that the first effect is observed.

#### *Intervention*

An intervention in Effect 1, symbolized by the "do" operator (e.g., do[e1]), blocks the causal influence of the common cause, which makes the cause and Effect 1 independent. (See Pearl, 2000, for more details on the "do" calculus.) Therefore, an intervention in E1 implies a modified causal model, which is captured by the following equations:

$$P(\mathrm{E2}|\mathrm{do}[\mathrm{e1}]) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}). \qquad (3)$$

$$P(\mathrm{E2}|\mathrm{do}[{\sim}\mathrm{e1}]) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}). \qquad (3')$$

As the two equations indicate, the probability of the second effect is causally independent of an intervention in the first effect. The probability is identical in both cases. According to Equation 3, the probability of the second effect can be computed as

$$P(\mathrm{e2}|\mathrm{do}[\mathrm{e1}]) = P(\mathrm{e2}|\mathrm{do}[{\sim}\mathrm{e1}]) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c}).$$

A comparison between Equations 2 and 3 shows the formal differences between observation and intervention. Whereas in the case of an intervention, Cause C is independent of the effect that is generated by the intervention, the cause is dependent on the observation of the effect in the seeing context.

### Causal-Chain Model

In the causal-chain model (Figure 2), it is assumed that event E1 influences event C, which causes effect E2. According to the Markov condition, Event C screens off event E1 from event E2. Accordingly, the joint probability of the three events can be factored using the following formula

$$P(\mathrm{E2,E1,C}) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|\mathrm{E1}) \times P(\mathrm{E1}). \qquad (4)$$

#### *Observation*

Given that E1 is observed to be present (E1 = e1) or absent (E1 = ~e1), the model implies

$$P(\mathrm{E2}|\mathrm{e1}) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|\mathrm{e1}). \qquad (5)$$

$$P(\mathrm{E2}|{\sim}\mathrm{e1}) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|{\sim}\mathrm{e1}). \qquad (5')$$

The probability of the final effect E2 being present given that the first cause in the chain (E1) is observed can therefore be calculated by

$$P(\mathrm{e2}|\mathrm{e1}) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}|\mathrm{e1}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c}|\mathrm{e1})$$

and

$$P(\mathrm{e2}|{\sim}\mathrm{e1}) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}|{\sim}\mathrm{e1}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c}|{\sim}\mathrm{e1}).$$

#### *Intervention*

If the event E1 is generated or inhibited by an intervention, it will influence the presence of both C and E2:

$$P(\mathrm{E2}|\mathrm{do}[\mathrm{e1}]) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|\mathrm{do}[\mathrm{e1}]) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|\mathrm{e1}). \qquad (6)$$

$$P(\mathrm{E2}|\mathrm{do}[{\sim}\mathrm{e1}]) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|\mathrm{do}[{\sim}\mathrm{e1}]) = P(\mathrm{E2}|\mathrm{C}) \times P(\mathrm{C}|{\sim}\mathrm{e1}). \qquad (6')$$

Comparing Equations 6 and 6′ with Equations 5 and 5′ shows that in a causal-chain model with three events, observation and intervention have identical implications. Using Equations 6 and 6′, the probability of the final effect E2 can be computed by

$$P(\mathrm{e2}|\mathrm{e1}) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}|\mathrm{e1}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c}|\mathrm{e1})$$

and

$$P(\mathrm{e2}|{\sim}\mathrm{e1}) = P(\mathrm{e2}|\mathrm{c}) \times P(\mathrm{c}|{\sim}\mathrm{e1}) + P(\mathrm{e2}|{\sim}\mathrm{c}) \times P({\sim}\mathrm{c}|{\sim}\mathrm{e1}).$$

### Comparison Between Common-Cause Model and Causal-Chain Model

Comparing Equations 6 and 6′ with Equations 3 and 3′ shows that the final effect E2 is dependent on the intervention in e1 in a causal-chain model but not in a common-cause model. The comparison of Equations 5 and 5′ with Equations 2 and 2′ shows that both models entail the same predictions for observations of event E1. However, it should be noted that there is an essential difference between the two models even in this case. Whereas $P(\mathrm{c}|\mathrm{e1})$ indicates the posterior probability of the cause in a common-cause model, it represents the probability of the intermediate effect conditional on its cause in a causal-chain model. Thus, even when the models imply the same numeric predictions for observations, they are generated by different mechanisms, as the formulas show.

Received April 15, 2004
Revision received September 1, 2004
Accepted September 2, 2004 ■