Meder, B., Hagmayer, Y., & Waldmann, M. R. (2005). Doing after Seeing. In B.G. Bara, L. Barsalou, & M. Bucciarelli (Eds.) *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1461-1466). Mahwah, NJ: Erlbaum.

Doing After Seeing

Björn Meder (bmeder@uni-goettingen.de)
York Hagmayer (york.hagmayer@bio.uni-goettingen.de)
Michael R. Waldmann (michael.waldmann@bio.uni-goettingen.de)

Department of Psychology, University of Göttingen, Gosslerstr. 14,
37073 Göttingen, Germany

Abstract

Causal knowledge serves two functions: it allows us to predict future events on the basis of observations and to plan actions. Although associative learning theories traditionally differentiate between learning based on observations (classical conditioning) and learning based on the outcomes of actions (instrumental conditioning), they fail to express the common basis of these two modes of accessing causal knowledge. In contrast, the theory of causal Bayes nets captures the distinction between observations (seeing) and interventions (doing), and provides mechanisms for predicting the outcomes of hypothetical interventions from observational data. In two experiments, in which participants acquired observational knowledge in a trial-by-trial learning procedure, the adequacy of causal Bayes nets as models of human learning was examined. To test the robustness of learners' competency, the experiments varied the temporal order in which the causal events were presented (predictive vs. diagnostic). The results support the theory of causal Bayes nets but also show that conflicting temporal information can impair performance.

Introduction

The ability of acquiring and using causal knowledge is a central competency necessary for explaining past events, predicting future events, and planning actions. But how do people infer the consequences of their actions when planning interventions in causal systems? In some cases they might have tried out the interventions on previous occasions so that they already know the potential outcomes of their actions. But what if only knowledge about observational relations between causal events is available? A tempting solution would be to equate observational knowledge with instrumental knowledge and proceed from there. Unfortunately this strategy will often lead to ineffective actions. For example, the status of a barometer is statistically related to the upcoming weather, but manipulating the barometer does not affect the weather. Barometer readings and weather are spuriously related due to their common cause, atmospheric pressure. As a consequence, observational predictions can capitalize on the spurious statistical relation, whereas instrumental predictions cannot. Effects do not change their causes; thus, manipulating the barometer does not affect its cause atmospheric pressure, and therefore has no causal influence on the weather.

The difference between observing (seeing) and intervening (doing) is compelling in the barometer example. Nevertheless, most theories of causal cognition do not distinguish between different modes of accessing identical causal knowledge. For example, associative theories of causal learning are sensitive to covariations but do not distinguish between causal and spurious relations. It is true that these theories distinguish between observational (classical conditioning) and interventional learning (instrumental conditioning), but, as the barometer example shows, they fail when predictions for instrumental actions have to be derived from observational learning.

Causal Bayes nets (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; Woodward, 2003) provide a formal account of causal representations and inference that allows it to derive precise predictions for hypothetical interventions from observational knowledge. The goal of the present experiments is to investigate whether people who have observed individual trials presenting the states of a complex causal model can later access their causal knowledge to derive observational and interventional predictions in a fashion anticipated by causal Bayes nets. To test the robustness of this competency we manipulated the temporal cues in the learning data (predictive learning from causes to effects vs. diagnostic learning from effects to causes).

Seeing vs. Doing in Causal Bayes Nets

The formal framework of causal Bayes nets uses directed acyclic graphs (DAGs) to represent causal relations between variables, and parameters to express the strength of these relations (e.g., conditional probabilities). A complete causal model therefore combines qualitative assumptions about the structure of the causal model with quantitative knowledge about the size of the parameters associated with these causal relations (e.g., base rates, causal strength, integration rules). An example is given in Figure 1. This causal model consists of four (binary) variables *A*, *B*, *C,* and *D*, in which *A* can cause *D* either via *B* or via *C*.

In the causal Bayes nets framework, the joint probability distribution of this model can be factored into:

$$p(A.B.C.D)=p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B.C)$$

This decomposition follows from applying the *causal Markov condition* (Spirtes et al., 1993; Pearl, 2000) to the causal model. The causal Markov condition (informally) states that the state of any variable X_j in the system is independent of all other variables (except for its causal descendants) conditional on the set of its direct causes, pa_x. For example, the causal Markov condition implies for the causal model shown in Figure 1 that variable *D* is independent of *A* conditional on its direct causes *B* and *C*. Relations of conditional dependence and independence are critical for deriving observational and interventional inferences.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2005). Doing after Seeing. In B.G. Bara, L. Barsalou, & M. Bucciarelli (Eds.) *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1461-1466). Mahwah, NJ: Erlbaum.

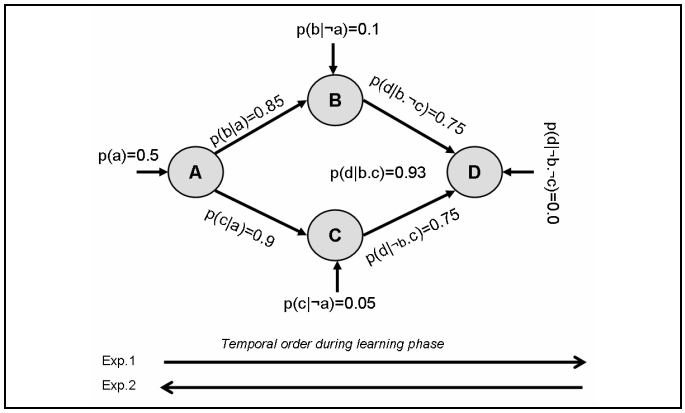

Figure 1: A parameterized causal model. Arrows indicate causal relations between variables; conditional probabilities encode the strength of these relations. All parameters were set except *p(d|b.c)* which is computed by a noisy-OR-gate.

Observational Predictions

Based on the structure of the causal model and its parameters, the probabilities implied by observations of the states of observed variables can be computed using probability calculus. For example, if variable *C* in the causal model shown in Figure 1 is observed, the probability of *A* being present can be computed using Bayes rule,

$$p(a|c)=p(c|a) \cdot p(a)/[p(c|a) \cdot p(a)+p(c|\neg a) \cdot p(\neg a)].$$

This computation models a diagnostic inference from *C* to its cause *A*. A more interesting example is the prediction of variable *D* from an observation of variable *C*. Obviously there is the direct causal link connecting *C* to *D*. But there is also a second causal pathway connecting *C* to *D* via *A* and *B*. If *C* is present, the probability of *A* increases, which in turn increases the probability of *B* leading to an increase of *D*. Pearl (2000) vividly calls such confounding pathways *backdoors*. Formally the probability for *D* given *C* can be calculated by:

$$p(d|c)=p(a|c)\cdot p(b|a)\cdot p(d|b.c)+p(a|c)\cdot p(\neg b|a)\cdot p(d|\neg b.c)+ p(\neg a|c)\cdot p(b|\neg a)\cdot p(d|b.c)+p(\neg a|c)\cdot p(\neg b|\neg a)\cdot p(d\neg b.c).$$

Interventional Predictions

The literature on causal Bayes nets has focused on ideal interventions in which the actions change the value of a variable independent of the state of its parents (for more precise characterizations of these interventions, see, for example, Woodward, 2003). For example, if we arbitrarily change the reading of the barometer our action renders the barometer independent of its usual cause, atmospheric pressure. Thus, interventions create independence, which can be expressed by removing the causal links between the variable that is targeted by the interventions and its parents. For example, if variable *C* in the causal model depicted in Figure 1 is set to a specific value through an intervention, the causal arrow from *A* to *C* can be eliminated. Following previous work of Spirtes et al. (1993), Pearl (2000) describes this process as 'graph surgery'. Interventional predictions should be based on this modified manipulated graph and not on the original graph. Because of graph surgery, interventions in contrast to observations do not provide diagnostic evidence for the causes of the manipulated variable.

To formalize the idea that a variable's state is not based on the 'natural course of events' but was determined by an external intervention, Pearl (2000) introduced the so-called '*Do-Operator'*, written as Do (•).The expression 'Do C=c' is read as "variable *C* is set to state *c* by means of an intervention". Formally, the Do-operator renders a variable independent of its direct causes, which is equivalent to deleting all causal links pointing towards the variable fixed by the intervention. Based on the modified causal model, the probabilities of the other events can be computed. For example, the probability of *A=a* given that *C* is caused by an intervention equals the base rate of *A=a* because the causal link connecting these two events was eliminated by the intervention,

$$p(a|Do(C=c)) = p(a|Do(C= \neg c)) = p(a).$$

In the same way the probability of *D=d* can be calculated using the modified causal model. Generating a value of *C* through an intervention closes the backdoor, hence no second causal pathway remains. Nevertheless the initial cause *A* may occur and influence *D* via *B*. Therefore the correct formula to calculate the probability of *D=d* given a (generative) intervention in *C* is:

$$p(d|Do(c))=p(a)\cdot p(b|a)\cdot p(d|b.c)+p(a)\cdot p(\neg b|a)\cdot p(d\neg b.c) +p(\neg a)\cdot p(b|\neg a)\cdot p(d|b.c)+p(\neg a)\cdot p(\neg b|\neg a)\cdot p(d|\neg b.c).$$

Note that variable *A* is no longer conditionalized on *C* in this formula. This implies that the interventional probability is necessarily lower than the observational probability, provided that *A* and *B* are positively related.

On the right hand side of the equations only parameters are involved which can be derived from observational learning. Thus, subsequent to observational learning of a causal model and of its parameters, the consequences of hypothetical interventions can be predicted without prior instrumental learning.

Psychological Evidence

The examples described above show that normatively diagnostic and predictive inferences differ depending on whether they are based on interventions or observations. The interesting question is whether people make these distinctions as well. Thus far, very few experiments have addressed this question (see Hagmayer, Sloman, Lagnado, & Waldmann, in press, for an overview).

Sloman and Lagnado (2005) have studied counterfactual inferences in given causal structures and have compared causal with logical arguments. For example, participants were given a causal chain model consisting of three events that were all described to be present. Participants were then requested to imagine that the intermediate event was either removed by an intervention or observed to be absent. In accordance with the predictions of causal Bayes nets participants inferred that the initial cause in the chain would be absent if the intermediate event was observed to be absent, but not if it was actively removed. Overall the

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2005). Doing after Seeing. In B.G. Bara, L. Barsalou, & M. Bucciarelli (Eds.) *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1461-1466). Mahwah, NJ: Erlbaum.

results of Sloman and Lagnado's experiments were consistent with causal Bayes nets. Because the focus of these studies was on comparing logical with causal reasoning, only qualitative reasoning based on the structure of causal models was investigated.

Waldmann and Hagmayer (2005) wondered whether a dissociation between seeing and doing can also be found in the realm of learning, and whether learners' inferences would be sensitive to the size of the parameters that were gleaned from the learning data. Participants in their experiments were given instructions about the structure of causal models and subsequently received a list of cases on a single page that could be used to estimate the parameters of the models. Participants were then requested to derive predictions for new hypothetical observations and hypothetical interventions. The results showed a surprising grasp of the differences between seeing and doing, which manifested itself in predictions that took into account the size of the parameters which were estimated on the basis of the learning data.

The present experiments move one step further in the realm of learning. In Waldmann and Hagmayer (2005) the parameters could be estimated on the basis of a list of cases which provided simultaneous information about the presence or absence of the variables. It can be argued that this task is still more a reasoning task than a learning task. The typical temporal characteristics of causal learning are better mirrored in trial-by-trial learning than in a highly processed list that lacks natural temporal cues. In fact, Shanks (1991) has hypothesized that induction on the basis of aggregated data is handled by different learning mechanisms than trial-by-trial learning. Thus, a demonstration of the competency to distinguish seeing and doing in the context of trial-by-trial learning would further weaken associative theories as models of causal induction.

Moreover, studying inferences based on trial-by-trial learning introduces cues to causal structures that might compete with the instructed causal models. The arrows within causal models express our natural intuition about the asymmetry of causes and effects: Causes generate effects but not vice versa. In the real world causal order is often signaled by the temporal order in which causes and effects are experienced. However, there are cases in which temporal order and causal order mismatch. For example, physicians often observe symptoms (i.e., effects) prior to learning about their causes. In these cases it is crucial that the temporal order of experiencing events is ignored as a cue to causality.

Trial-by-trial learning which presents learning events in a temporal order allows it to test the impact of temporal cues on causal induction. A number of experiments designed to test causal-model theory (Waldmann, 1996) have pitted temporal order against causal order. In these experiments it could be shown that learners are capable of learning causal models regardless of whether temporal order matches or mismatches causal order. However, this competency breaks down when complexity is increased (Waldmann & Walker, 2005). Moreover, the competency was only tested with test questions that requested observational predictions. Interventional questions are more complex because they require a stage of model manipulation (e.g., graph surgery) prior to using the manipulated model for the predictions. To test whether learners distinguish between seeing and doing even in conditions in which the temporal order of learning events mismatches causal order, we varied learning order in the experiments.

In Experiment 1 the temporal order during each trial conformed to the causal order of the events in the causal model (see Fig. 1). Information about *A* was given first, followed by information about the presence or absence of *B* and *C*, and finally information about *D* was provided. In Experiment 2 the temporal order was reversed and no longer matched the causal order of events. In this experiment it is necessary to suppress temporal cues and estimate the parameters on the basis of data that is presented in the reverse order (see Waldmann & Martignon, 1998). For example, in the causal model we have used in the experiments (Fig. 1), the final effect *D* is dependent on patterns of its two causes *B* and *C*. When temporal order is reversed in Experiment 2, participants observed the probability of *B* and *C* given *D* and the probability of *A* given *B* and *C* but have to infer the probabilities of *B* and *C* conditional upon *A*, and of *D* conditional upon *B* and *C*. Note that the patterns of covariation are nevertheless exactly the same across Experiments 1 and 2, that is, participants' judgments are based on the very same data. Only the order in which information about the events was given is manipulated across the two experiments. Although we expect that, consistent with causal-model theory, learners will attempt to correctly learn the causal model regardless of learning order, and will differentiate between seeing and doing, this competency might be marred by performance deficits caused by the complexity of the task (see also Waldmann & Walker, 2005).

A further novel aspect of this study is the presentation of a causal model that contains two parallel pathways that represent mutual confounds (i.e., backdoors). One additional goal of the experiments was to test whether learners are sensitive to the fact that interventions and observations differ with respect to the way the second confounding pathway needs to be taken into account.

Experiment 1

The goal of the first experiment was to investigate whether learners differentiate between seeing and doing after a trial-by-trial learning phase in which learning order corresponds to causal order. Twenty-four students from the University of Göttingen participated. The model underlying the learning data and its parameterization are shown in Figure 1. After the observational learning phase participants were asked to imagine new cases in which either variable *C* was observed to be present or absent, or *C* was generated or eliminated by an intervention. All participants had to estimate the probability of *A* and *D* in each of these four cases (i.e., eight questions).

Learning Phase. The variables of the causal model were introduced as four fictitious chemical substances causally interacting in wine casks. Participants were told that substance *A* causes the generation of substances *B* and *C*,

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2005). Doing after Seeing. In B.G. Bara, L. Barsalou, & M. Bucciarelli (Eds.) *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1461-1466). Mahwah, NJ: Erlbaum.

which then can independently cause substance *D* (see Fig. 1). It was also pointed out that the causal relations are probabilistic. In addition, participants were shown the graph of the causal model. They were instructed to learn the strength of the causal relations from the learning data. The kind of questions they would have to answer after the learning phase was not mentioned until the test phase. The learning phase consisted of 40 trials which implemented the probabilities shown in the causal graph in Figure 1. The trials presented information on a computer screen about the states of the variables. The temporal order corresponded to causal order in Experiment 1. Thus, first information about variable *A* was presented, then, simultaneously, variables *B* and *C* were shown, and finally information about event *D* was given.

Test Phase. The learning phase was followed by a test phase in which participants had to answer eight questions. The questions first stated the current status of variable *C* (present vs. absent) and then asked about *A* (i.e., the cause of *C*) or *D* (i.e., the effect of *C*). Thus, one question was diagnostic, the other predictive. For the observational predictions, participants were instructed to imagine observing substance *C* in 40 previously unseen wine casks and to estimate the number of casks in which substance *A* would also be found, (i.e., they estimated $p(a|c)$ in a frequency format). Participants were also asked to estimate the conditional frequency of *A* for 40 casks in which *C* was observed to be absent ($p(a|\neg c)$). Two further questions referred to interventions. These questions asked learners to imagine that substance *C* was added to 40 casks ($p(a|Do(c))$), or that *C* was inhibited in 40 casks ($p(a|Do(\neg c))$). The same set of questions was asked about *D*, the effect of *C*. Thus, participants estimated the number of casks in which *D* would be found (i.e., $p(d|c)$, $p(d|\neg c)$, $p(d|Do(c))$, $p(d|Do(\neg c))$. Interventional and observational questions were blocked; the order of blocks was counterbalanced.

Causal Bayes Net Predictions. For the diagnostic inferences, the crucial test for assessing whether participants differentiated between observations and interventions concerns the comparison of the two observational probabilities $p(a|c)$ versus $p(a|\neg c)$ to the two corresponding interventional probabilities $p(a|Do(c))$ versus $p(a|Do(\neg c))$. Whereas observing $C{=}c$ or $C{=}\neg c$ is diagnostic evidence for the state of *A*, generating $C{=}c$ or $C{=}\neg c$ by an intervention is not diagnostic for *A*. Therefore, the observational probabilities should differ, whereas the interventional probabilities should stay at a constant level. According to causal Bayes nets, *A* should be expected to occur with the probability that corresponds to its base rate.

The predictive inferences are more complicated because the second causal pathway generating *D* has to be taken into account. Participants should consider both *C*'s direct causal relation to *D* but also the alternative path $A \rightarrow B \rightarrow D$. The observation of *C* opens the backdoor to the second pathway (i.e., the presence of *C* indicates that *A* is likely to be present). Therefore, participants should infer that *D* is more likely to be present if *C* is observed than when it is absent. In contrast, if *C* is manipulated by an intervention the backdoor is closed because the link between *A* and *C* needs to be removed. However, *D* is still more likely when *C* is generated than when it is inhibited because of *C*'s direct causal influence. The difference, however, should be smaller for the interventional than for the observational probabilities.

A further test of sensitivity to the difference between seeing and doing is provided by comparing the estimated probabilities of *D* given observations of *C* or interventions in *C*. In the section about causal Bayes nets it was mentioned that the probability of *D* is higher if *C* is observed than when it is generated by an intervention. The parameters of the presented causal model imply that the interventional probability of observing *C* ($p(d|c)$) is only slightly higher than the probability of generating *C* ($p(d|Do(c))$. But the probability of *D* is considerably higher when *C* is prevented $p(d|Do(\neg c))$ than when it is observed to be absent ($p(d|\neg c)$).

Results and Discussion

*Diagnostic inferences***.** The results for the diagnostic test questions are shown in Table 1 along with the predictions

Table 1: Responses to diagnostic inference questions in Experiment 1 (N=24) (Numbers indicate means of conditional frequency estimates for 40 cases.)

	Observation		**Intervention**	
	p(a|c)	***p(a|¬c)***	***p(a|Do(c))***	***p(a|Do(¬c))***
Causal Bayes net predictions	*38*	*4*	*20*	*20*
Mean	30.50	17.08	25.54	27.25
SD	7.56	10.37	10.57	8.59

derived from causal Bayes nets. The pattern of estimated conditional frequencies qualitatively conformed to the pattern of the predicted values. As anticipated by causal Bayes nets, participants gave different estimates for the two observational probabilities but judged the interventional probabilities to be at the same level. An analysis of variance with the factors 'intervention vs. observation' and 'presence vs. absence of *C*' as within-subjects factors yielded a significant interaction, $F(1,23)=23.78$, $p<0.001$, $MSE=57.75$. As predicted by the causal Bayes nets framework, there was no difference between the interventional questions, $F<1$, but a significant difference between the observational questions, $F(1,23)=35.51$, $p<0.001$, $MSE=59.17$. Although participants' estimates did not perfectly match the normative causal Bayes net predictions, the results provide evidence for participants' sensitivity to the difference between seeing and doing in diagnostic judgments.

Predictive inferences. The results for the predictive questions concerning the probability of *D* are shown in Table 2. This type of inference is more complicated than the diagnostic inference in the chosen causal model. Whereas the latter only requires considering the direct causal relation between *A* and *C* (with the rest of the causal model being

irrelevant for this task), the inference concerning variable D requires taking into account the complete model. In particular, the alternative confounding pathway $A \rightarrow B \rightarrow D$ needs to be considered.

As can be seen in Table 2, participants were surprisingly sensitive to the second confounding causal pathway: As predicted by causal Bayes nets, the difference between responses to the observational questions proved larger than the difference between the responses to the interventional questions. An analysis of variance with 'intervention vs. observation' and 'presence vs. absence of *C*' as within-subjects factors yielded a significant interaction, $F(1,23)=8.73$, $p<0.01$, $MSE=54.65$. In accordance with the parameterization of the causal model, there was only a slight, non-significant difference between $p(d|c)$ and $p(d|Do(c))$, $F(1,23)=1.0$, $p=0.33$. This is important as there might have been a general tendency to answer interventional questions differently from observational questions. The crucial test of the predictions of causal Bayes nets is provided by the comparison between $p(d|\neg c)$ and $p(d|Do(\neg c))$. Participants judged the probability of the occurrence of *D* to be significantly higher when *C* was prevented by an intervention than when it was merely observed to be absent, $F(1,23)=9.57$, $p<0.01$, $MSE=57.83$.

The results demonstrate a remarkable grasp of the difference between intervening and observing after trial-by-trial learning of a causal model. Both diagnostic and predictive inference proved sensitive to the distinction between seeing and doing. The experiment also provides evidence for learners' sensitivity to the implications of alternative confounding pathways (i.e., backdoors) for observations and interventions. Although the statistical patterns correspond to the predictions of causal Bayes nets, the estimates were not perfect, of course. Specifically, learners had difficulties with correctly assessing cases in which events were observed to be absent (see also Waldmann & Hagmayer, 2005). The complexity of the model and the limited number of learning trials might have contributed to the imperfections. Nevertheless, the competency of the participants was remarkable and provides clear evidence against traditional learning theories that fail to account for complex causal model learning.

Table 2: Responses to predictive inference questions in Experiment 1 (*N*=24) (Numbers indicate means of conditional frequency estimates for 40 cases.)

	Observation		**Intervention**	
	p(d|c)	***p(d|¬c)***	***p(d|Do(c))***	***p(d|Do(¬c))***
Causal Bayes net predictions	*36*	*5*	*33*	*14*
Mean	29.67	14.79	27.54	21.58
SD	10.04	11.56	11.64	12.55

Experiment 2

The main goal of Experiment 2 was to test whether people learn and access causal models adequately when the learning order does not match causal order. Previous research has shown that people can make correct predictions after diagnostic learning (effects presented prior to their causes) but this competency was only tested with observational test questions and displayed itself only with causal models that were less complex than the causal model used in the present experiments (see Waldmann & Walker, 2005). Experiment 2 used the same experimental design, cover story, and instructions as Experiment 1. Thus, participants received instructions about the causal model displayed in Figure 1. Again 24 participants from the University of Göttingen participated. In contrast to Experiment 1, the temporal order of learning events did not match their causal order (i.e., diagnostic learning). In each trial, participants were first informed about the status of effect *D*, then simultaneously about the mediating variables *B* and *C*, and finally about the initial cause *A*. Participants had to mentally reverse the observed statistical relations to correctly estimate the parameters of the causal model. As in Experiment 1, participants were requested to estimate the conditional frequencies of *A* and *D* given observations of or interventions in *C*.

Results and Discussion. Tables 3 and 4 show the means of the conditional frequency estimates along with the predictions derived from causal Bayes nets.

Table 3: Responses to diagnostic inference questions in Experiment 2 (*N*=24) (Numbers indicate means of conditional frequency estimates for 40 cases.)

	Observation		**Intervention**	
	p(a|c)	***p(a|¬c)***	***p(a|Do(c))***	***p(a|Do(¬c))***
Causal Bayes net predictions	*38*	*4*	*20*	*20*
Mean	33.46	15.38	25.50	22.42
SD	8.59	11.20	11.16	10.31

Diagnostic inferences. The mean estimates for the conditional frequency of *A* given *C* closely resemble the ones in Experiment 1. Again the general pattern corresponds to the predictions of causal Bayes nets. An analysis of variance with the factors 'intervention vs. observation' and 'presence vs. absence of *C*' as within-subjects factors again yielded a significant interaction, $F(1,23)=28.15$, $p<0.001$, $MSE=47.96$. The difference between the two estimated observational probabilities proved considerably larger than the one between the two interventional probabilities. Again participants judged the probability of *A* to be at a similar level regardless of whether *C* was generated or prevented. Specifically, the observational questions differ significantly, $F(1,23)=63.88$, $p<0.001$, $MSE=61.43$ while there is no difference between the interventional questions, $F(1,23)=1.52$, $p=0.23$. The diagnostic inferences show a remarkable grasp of the difference between seeing and doing despite the added complexity of the diagnostic learning procedure. In this task, participants proved capable of ignoring the misleading temporal cue of the learning procedure.

Predictive inferences. As in Experiment 1, participants were asked to estimate the probability of *D* when *C* was observed

or manipulated by an external intervention. However, in contrast to Experiment 1, the estimates considerably deviated from the causal Bayes net predictions (see Table 4). An analysis of variance yielded no significant interaction. In addition, the observational probability estimate of $p(d|\neg c)$ did not differ significantly from the corresponding interventional probability estimate of $p(d|Do(\neg c))$, $F<1$. Thus, there was no evidence that participants correctly differentiated between seeing and doing in the predictive task. Probably the increased complexity caused by the misleading temporal cue and the complicated inference, which required taking into account a secondary confounding pathway (i.e., backdoor), exceeded the information processing capacity of learners

Table 4: Responses to predictive inference questions in Experiment 2 (N=24) (Numbers indicate means of conditional frequency estimates for 40 cases.)

	Observation		**Intervention**	
	p(d|c)	***p(d|¬c)***	***p(d|Do(c))***	***p(d|Do(¬c))***
Causal Bayes net predictions	*36*	*5*	*33*	*14*
Mean	30.54	18.33	29.71	20.33
SD	10.10	13.26	10.90	11.74

General Discussion

Taken together, the results of the two experiments provide convincing evidence that learners are capable of distinguishing between observations and interventions even in a more naturalistic learning environment. These findings contradict traditional associative learning theories, which fail to model causal-model learning and which are incapable of deriving correct predictions for actions after purely observational learning. The present experiments also demonstrate a surprising grasp of the implications of confounding pathways. Thus, the experiments strongly support causal-model theory and causal Bayes nets as theoretical accounts of causal induction.

However, Experiment 2 also shows that the competence of learners only displays itself when the complexity of the task does not exceed learners' information processing capacity (see also Waldmann & Walker, 2005). A popular strategy to deal with such impairments is to postulate two systems, a reasoning component that handles summarized data, and an associative learning component that is specialized for trial-by-trial learning (Price & Yates, 1995; Shanks, 1991). Although Experiment 1, which used a trial-by-trial learning procedure, already weakens this account, it might still be speculated that learners fell back on an associative mode in Experiment 2 in which the learning task was more complex. However, the data of Experiment 2 are inconsistent with this theory. Learners were not generally impaired, only the predictive inferences were affected. The less complex diagnostic inferences showed a remarkable grasp of the differences between seeing and doing despite the misleading temporal cue. Only the more complex predictive inferences were negatively affected.

We believe that the reason for this difference is located in the parameter estimation processes. Participants in Experiment 2 observed the probability of B and C given D but had to infer the probability of D given B and C as a parameter of the causal model. Such an inversion is complicated and demanding. Therefore, the learning process may have led to inadequate estimates of the model's parameters. The diagnostic questions could be correctly answered by recognizing that interventions render the manipulated variables independent of their actual causes, which implies that solely the base rate $p(A)$ needs to be accessed for giving a correct response. In contrast, the predictive questions can only be answered correctly if the parameters of the full causal model are correctly estimated, and if the model is correctly altered for the intervention questions. Thus, if the parameters are not acquired correctly during learning the inferences are likely to be wrong. This account explains the deficits shown in Experiment 2 using a causal Bayes net analysis of the task. Future research will have to develop psychological models of learning that integrate competence and performance.

References

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (in press). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.): *Causal learning: Psychology, philosophy, and computation.* Oxford: Oxford University Press.

Pearl, J. (2000) *Causality*. 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, and Cognition*, *21*, 1639-1655.

Shanks, D. R. (1991). On similarities between causal judgments in experienced and described situations. *Psychological Science, 5*, 341-350.

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

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

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

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. (2005). Competence and performance in causal learning. *Learning & Behavior.*

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