Inferences about unobserved causes in human contingency learning

PLEASE SCROLL DOWN FOR ARTICLE

Psychology Press
Taylor & Francis Group

Inferences about unobserved causes in human contingency learning

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

Estimates of the causal efficacy of an event need to take into account the possible presence and influence of other unobserved causes that might have contributed to the occurrence of the effect. Current theoretical approaches deal differently with this problem. Associative theories assume that at least one unobserved cause is always present. In contrast, causal Bayes net theories (including Power PC theory) hypothesize that unobserved causes may be present or absent. These theories generally assume independence of different causes of the same event, which greatly simplifies modelling learning and inference. In two experiments participants were requested to learn about the causal relation between a single cause and an effect by observing their co-occurrence (Experiment 1) or by actively intervening in the cause (Experiment 2). Participants' assumptions about the presence of an unobserved cause were assessed either after each learning trial or at the end of the learning phase. The results show an interesting dissociation. Whereas there was a tendency to assume interdependence of the causes in the online judgements during learning, the final judgements tended to be more in the direction of an independence assumption. Possible explanations and implications of these findings are discussed.

Most events have many causes, most of which we do not know or cannot observe. Nevertheless, despite the fact that the observed effects are also influenced by unobserved events, we are capable of learning about the causal efficacy of observed causes. A large body of theoretical and empirical knowledge has been accumulated on how we make causal strength assessments in such situations (Cheng, 1997; Shanks, 1987; Shanks, Holyoak, & Medin, 1996; Waldmann & Hagmayer, 2001; Wasserman, Elek, Chatlosh, & Baker, 1993; White, 2003). However, the question

has largely been neglected in the literature as to what we can learn about the causes that we cannot observe (but see Kushnir, Gopnik, Schulz, & Danks, 2003; Luhmann & Ahn, 2003). For example, in the early days of AIDS nobody knew about the virus and its lethal capacity. The known observable causes included minor infections, which apparently led to serious, live threatening conditions, such as pneumonia. However, the seriousness of the disease in light of the rather harmless causes led to the search for additional unobserved causes. Eventually the

HIV virus was discovered, and theories were developed that traced the complex causal pathways underlying the disease.

Although most theories of causal learning did not directly address inferences about unobserved causes, most theories took the possibility of such causal factors into account. The question of how to assess causal strength when there are unobserved causes has challenged normative theories of causality and psychological theories of causal reasoning for some time. A number of different accounts have been proposed analysing how the causal strength of an observed factor can be accurately estimated if certain assumptions are made about potential unobserved causes. In this article we study more directly what people infer about unobserved causes during learning. First we give a brief overview of how associative and causal Bayes net theories handle unobserved causes. Based on these theories we derive predictions about the inferences that people should make about the presence and causal influence of unobserved causes. In the second part of the article we present two experiments in which we assessed the assumptions of learners about the strength and probability of unobserved causes. In the final section we discuss potential theoretical implications of these findings.

Theoretical accounts of unobserved causes

Unobserved events can be linked in complex networks and can simultaneously cause several observed events, leading to confoundings (see Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). In the present research we focus on a simple causal situation consisting of a single observable cause C and one possible unobserved cause A, both influencing a joint observable effect E. Figure 1 illustrates the causal model. The two observable events C and E are statistically related. There are three questions one might ask in this kind of situation: How does C influence E? How are C and A related to each other? How does A influence E?
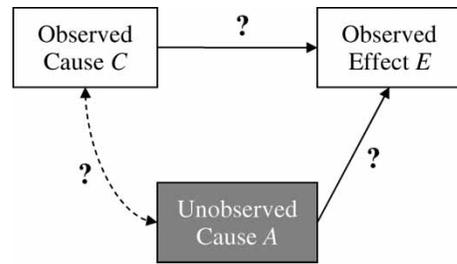
Figure 1. *Causal model of a situation in which an observable cause* C *and an unobserved cause* A *both influence an observable effect* E. *The question marks indicate the relations that participants were requested to assess.*

Associative theories

Associative theories, such as the Rescorla–Wagner theory (Rescorla & Wagner, 1972), would model this task as learning about the association between a cue representing the observed cause and an outcome representing the effect. Along with the observable cause cue a second background (or context) cue would be part of the model. This background cue is assumed to be always present and to represent all other factors that might also generate the outcome. Thus, the background cue would play the role of representing the unobserved cause A in the outlined causal model. According to the Rescorla–Wagner rule, only weights of cues that are present in the learning trial are being updated. Therefore, the permanently present background cue will generally compete with the cause cue in cases in which the cause cue is present. If the outcome is also present, the associative weights of both cues will be raised; if the outcome is absent, the weights will be lowered. However, in cases in which the cause cue is absent, only the weight of the background cue will be altered. At the asymptote of learning the associative weight of the observed cause will equal the contingency of the cause cue and the outcome, which is defined as the numeric difference between the probabilities of the effect in the presence and the absence of the observed cause—that is, $\Delta P = P(e|c) - P(e|\sim c)$; (see Cheng, 1997; Danks, 2003). The associative weight of the background cue will correspond to the probability of the outcome in the absence of

the cause cue. Thus, the more often the outcome (= effect) occurs on its own, the higher the associative weight of the background cue will be.

Consequently associative theories would predict that participants assume an unobserved cause to be always present. In addition, these theories predict that at the asymptote the unobserved cause's strength would match the probability of the effect in the observable cause's absence, which corresponds to the associative weight linking the context cue and the effect, and that the estimated causal influence of the observable cause corresponds to the observed contingency.

Causal Bayes nets

Causal Bayes net theories of causal learning offer accounts for a large variety of problems—for example, for hypothesis testing, model generation, and parameter estimation (see Danks, in press; Glymour, 2001; Griffiths & Tenenbaum, 2005; Spirtes et al., 1993; Tenenbaum & Griffiths, 2003). In fact, causal Bayes net theories are a rather heterogeneous group of theories. While all theories agree on the representation of causal structures as acyclic graphs capturing relations of conditional dependence among the events, they assume different processes of learning and inference. Whereas some theories rely on Bayesian inferences involving the application of Bayes' theorem (e.g., Tenenbaum & Griffiths, 2003), other theories do not (e.g., Cheng's, 1997, Power PC theory; see below).

All causal Bayes net theories agree on the assumption that causal learning and reasoning are based on structured causal models (i.e., causal Bayes nets) of a domain. Causal Bayes nets consist of graphs of causal models and parameters representing base rates, causal strength, and causal interactions. The graph depicted in Figure 1 depicts a common-effect structure in which two causes—one observable and one unobserved—affect a single effect. The graphs specify relations of conditional dependence and independence amongst the variables within the model. For example, a common-effect model such as the one depicted in Figure 1 represents the assumption that the observed cause and the unobserved cause

are independent of each other, that the two causes are related to the effect, and that the two causes are negatively related conditional upon the effect—that is, if the effect and one of the causes is observed the presence of the second cause is less likely than its presence (i.e., explaining away). Thus, causal Bayes net theories in general assume that different causes of a common effect are independent of each other, as long as there is no further causal factor that affects the two causes and thereby creates dependence between them.

Power PC theory. Cheng's (1997) Power PC analysis of the causal impact of a single cause can be viewed as a special case of a causal Bayes net in which two causes independently influence a joint common effect (see Glymour, 2001). Based on this theory some predictions for the probability and causal power of an unobserved cause can be inferred. The theory presupposes that the occurrence of the effect E is a consequence of the causal powers of the observed cause C (p_c), of an unobserved cause A (p_a), and of their base rates $P(c)$ and $P(a)$. Formally the probability of the effect equals the sum of the base rates of the two causes multiplied by their causal power minus the intersection of the causes multiplied by both causal powers:

$$P(e) = P(c) \cdot p_c + P(a) \cdot p_a - P(c) \cdot P(a) \cdot p_c \cdot p_a.$$

Therefore the probability of the effect E conditional upon an observation of cause C— that is, $P(c) = 1$—is

$$P(e|c) = p_c + P(a|c) \cdot p_a - P(a|c) \cdot p_c \cdot p_a \quad 1$$

The probability of the effect given that the observable cause C is absent—that is, $P(c) = 0$—is

$$P(e|\sim c) = P(a|\sim c) \cdot p_a \quad 2$$

Equations 1 and 2 hold irrespective of whether the observed cause C and the unobserved cause A are dependent, $P(a|c) \neq P(a|\sim c)$, or independent, $P(a|c) = P(a|\sim c)$. Both Equation 1 and Equation 2

have at least two parameters that cannot be directly estimated on the basis of the observed relation between C and E. Thus, neither the probability nor the causal power of the unobserved cause can be directly inferred from the observable data.

Power PC theory makes the additional assumption that different causes of the same effect are independent of each other—that is, the theory assumes that the observed and the unobserved causes are not statistically related, $P(a|c) = P(a|\sim c) = P(a)$. This assumption allows it to derive estimates for the causal power of the observable causes even if there are other unknown causes generating the same effect. Based on this assumption Equations 1 and 2 can be simplified into

$$P(e|c) = p_c + P(a) \cdot p_a - P(a) \cdot p_a \cdot p_c \qquad 1'$$

The probability of the effect given that the observable cause C is absent—that is, $P(c) = 0$—is

$$P(e|\sim c) = P(a) \cdot p_a \qquad 2'$$

Inserting $2'$ into $1'$ yields

$$P(e|c) = p_c + P(e|\sim c) - P(e|\sim c) \cdot p_c \qquad 1''$$

which can be rearranged into the well-known formula for generative causal power

$$p_c = \frac{P(e|c) - P(e|\sim c)}{(1 - P(e|\sim c))}. \qquad \text{3; power formula}$$

What inferences can be made about the unobserved cause? The independence assumption of Power PC theory implies that the probability of the unobserved cause stays the same regardless of whether the observed cause has occurred or not. Moreover, Equation $2'$ implies that the causal power of the unobserved cause and its probability have to be at least as big as the probability of the effect in the absence of the observed cause, $p_a \geq P(e|\sim c)$ and $P(a) \geq P(e|\sim c)$. Equations $1'$ and $2'$ provide additional constraints for the two unknown parameters. Equation $2'$ requires that the product of the estimated probability and the

causal power of the unobserved cause has to equal the observable probability of the effect in the observable cause's absence. In sum, Power PC theory like other causal Bayes net theories assumes independence between causes and provides constraints for the probability and causal power of unobserved causes. This assumption is necessary to make precise inferences about the power of observed variables. However, these theories may also drop this assumption if it seems unwarranted. In the next section we give some examples for inferences that are possible without assuming independence.

Inferring unobserved causes

In this section we present two formal Bayesian analyses of potential inferences about an unobserved cause, which do not rely on the assumption of independence. The first analysis is based on probability calculus and concerns inferences about the probability and strength of a unobserved cause based on a set of observed data. The second analysis deals with inferences based on single observations of a given cause and an effect. The two analyses show that observable causal relations normatively allow some inferences to be drawn about unobserved causes, even if no independence among the causes is assumed. Readers not interested in the formal details may prefer to skip this section. The results of the analyses are summarized in the next section.

The first analysis is based on the causal model depicted in Figure 1. Two causes, C and A, one of which is not observable (A), are assumed to influence a joint effect E. Based on the chain rule of probability theory, the joint probabilities of the three events can be factorized as follows:

$$P(E, A, C) = P(E|A, C)^*P(A, C)$$

While the joint probability of the two observable events $P(E, C)$ can be estimated on the basis of the observable events, neither the probability of the effect conditional on the pattern of the observed and unobserved causes, $P(E|A,C)$, nor the joint probability of the observed and

unobserved cause, $P(A, C)$, can be estimated. Thus, additional assumptions have to be made. There is one popular assumption that is frequently made with causal Bayes nets (Jensen, 1997; Pearl, 1988), which is the so called noisy-or assumption. This assumption informally states (a) that each cause deterministically causes the effect unless its impact is intercepted by an inhibitor, and (b) that the inhibitors are independent of each other. Note that it is not assumed that the causes are independent. For example, if there are two causes of a common effect, and both are present, then either the first cause generates the effect or the second cause generates the effect or both do. Thus, the effect will only fail to occur if both causes are inhibited. Formally

$$P(e|a.c) = 1 - (1 - P(e|\sim a.c)) \cdot (1 - P(e|a.\sim c))$$

Based on the noisy-or rule the joint probability of cause and effect co-occurring can be calculated by

$$\begin{aligned} P(e.c) &= P(e|a.c) \cdot P(a|c) \cdot P(c) \\ &+ P(e|\sim a.c) \cdot P(\sim a|c) \cdot P(c) \\ &= \{1 - [1 - P(e|\sim a.c)] \cdot [1 - P(e|a.\sim c)]\} \\ &\cdot P(a|c) \cdot P(c) + P(e|\sim a.c) \cdot P(\sim a|c) \cdot P(c) \end{aligned}$$

Conditionalized on the presence of the observable cause C, the following equation results:

$$\begin{aligned} P(e|c) &= \{[1 - [1 - P(e|\sim a.c)] \\ &\cdot [1 - P(e|a.\sim c)]\} \cdot P(a|c) \\ &+ P(e|\sim a.c) \cdot P(\sim a|c) \end{aligned} \qquad 4$$

The probability of the effect occurring on its own can also be calculated:

$$\begin{aligned} P(e.\sim c) &= P(e|a.\sim c) \cdot P(a|\sim c) \cdot P(\sim c) \\ &+ P(e|\sim a.\sim c) \cdot P(\sim a|\sim c) \cdot P(\sim c) \end{aligned}$$

According to the noisy-or assumption absent causes never lead to the effect—that is, $P(e|\sim$

$a.\sim c) = 0$. Therefore

$$P(e.\sim c) = P(e|a.\sim c) \cdot P(a|\sim c) \cdot P(\sim c)$$

Conditional on the absence of the observable cause C, this equation can be further simplified into

$$P(e|\sim c) = P(e|a.\sim c) \cdot P(a|\sim c) \qquad 5$$

It should be noted that thus far no assumption about the statistical relation between the observable cause and the unobserved cause has been made. However, Equations 4 and 5 also show that no precise inferences about the probability and causal strength of the unobserved cause can be drawn on the basis of the observable events alone. There are at least two unknown parameters in each of these equations, which refer to the conditional probabilities that involve the unobservable cause. Nevertheless, Equations 4 and 5 do provide constraints for the possible causal strength and probability of the unobserved cause. In particular the probability of the effect in the absence of the observed cause provides a rather simple constraint. As Equation 5 shows, this probability represents a lower boundary for the probability of the unobserved cause in the observable cause's absence, $P(a|\sim c) \geq P(e|\sim c)$, and for the probability that this cause generates the observable effect on its own, which corresponds to its causal power, $P(e|a.\sim c) \geq P(e|\sim c)$). Thus, even if no precise inferences about the parameters of unobserved causes can be drawn this analysis shows that there are numeric boundaries for plausible inferences. The more often the effect occurs on its own, the more often the unobserved cause has probably been present and the stronger its causal impact upon the visible effect will be.

The second analysis shows that at least approximate inferences about the presence of the unobserved cause A can be drawn by observing the presence or absence of the cause C and effect E on individual trials. There are four possible patterns of events that might be observed: cause C and effect E both being present ($e.c$), cause C being present without the effect ($\sim e.c$), the effect being present without the observable cause

($e.\sim c$), and both events being absent ($\sim e.\sim c$). From a Bayesian perspective the question is whether it is more likely for each of these four possible observations that an unobserved cause A is present (a) or that it is absent ($\sim a$). Using Bayes' theorem the probability of the unobserved cause being present can be calculated as

$$P(A|E,\,C) = \frac{P(E|A,\,C)\cdot P(A|C)}{P(E|C)}.$$

Given that the presence of both cause C and the effect E are being observed, the posterior odds for an unobserved cause to be present can be calculated by

$$\frac{P(a|e.c)}{P(\sim a|e.c)} = \frac{P(e|a.c)}{P(e|\sim a.c)}\cdot\frac{P(a|c)}{P(\sim a|c)} \qquad 6$$

(i.e., posterior odds = likelihood ratio ·

prior odds)

Let us first consider the likelihood ratio: The relevant question is whether the probability of the effect is higher given both the unobserved and the observed cause are present or higher given only the observable cause is present—that is, whether $P(e|a.c) > P(e|\sim a.c)$. If the unobserved cause is assumed to have a positive causal influence upon the effect, then the effect is more likely if both causes are present than if only one cause is present, $P(e|a.c) > P(e|\sim a.c)$. Therefore, an observation of both the cause and the effect makes the presence of an unobserved cause more likely. Nevertheless, no definite prediction is possible without taking into account the prior odds. As long as $P(a|c)$ is equal or bigger than $P(\sim a|c)$, the prior odds are larger than or equal to one, which implies that the presence of the unobserved cause is more likely than its absence. This means that as long as we have reason to assume that the presence of an unobserved cause is a priori equally likely or more likely than its absence, its presence should be predicted if a cause and its effect are both being observed.

We now turn to the second possible pattern of events. If only the cause is observed without the effect then the posterior odds of the unobserved cause are given by

$$\frac{P(a|\sim e.c)}{P(\sim a|\sim e.c)} = \frac{P(\sim e|a.c)}{P(\sim e|\sim a.c)}\cdot\frac{P(a|c)}{P(\sim a|c)} \qquad 7$$

If both causes have a positive causal influence then $P(e|a.c) > P(e|\sim a.c)$ and therefore $P(\sim e|a.c) < P(\sim e|\sim a.c)$. Thus, an observation of a cause without its effect makes the presence of an unobserved cause less likely. Again a guess about the prior odds is needed to arrive at a definite judgement.

However, it is important to note that the prior odds are identical in all cases in which the cause is observed. Regardless of whether $e.c$ or $\sim e.c$ is being observed, the prior odds are $P(a|c)\,/\,P(\sim a|c)$ (see Equations 6 and 7). This implies that the posterior odds of the unobserved cause are higher in cases in which both cause and effect are observed than in cases in which only the cause is observed no matter what the prior odds are. Thus, even if no definite estimates are possible without specific assumptions about the prior probabilities of the unobserved cause, the posterior odds should differ.

If the effect is observed without the observable cause, the posterior odds are calculated as

$$\frac{P(a|e.\sim c)}{P(\sim a|e.\sim c)} = \frac{P(e|a.\sim c)}{P(e|\sim a.\sim c)}\cdot\frac{P(a|\sim c)}{P(\sim a|\sim c)} \qquad 8$$

Thus, as long as the unobserved cause has a spositive influence, $P(e|a.\sim c) > P(e|\sim a.\sim c)$, this observation makes the presence of the unobserved cause more likely. However, an even more distinct prediction is possible based on the often-stated causal principle that nothing happens without a cause ("nihil fit sine causa", Audi, 1995). Based on this principle, $P(e|\sim a.\sim c)$ should equal zero, which implies that the unobserved cause has to be present for sure. Thus, an observation of an effect without any visible cause makes the presence of an unobserved cause necessary.

Finally, if neither the cause nor its effect is present, the posterior odds of the unobserved

cause are

$$\frac{P(a|\sim e.\sim c)}{P(\sim a|\sim e.\sim c)} = \frac{P(\sim e|a.\sim c)}{P(\sim e|\sim a.\sim c)}$$

$$\cdot \frac{P(a|\sim c)}{P(\sim a|\sim c)} \qquad 9$$

If the effect never happens without any cause then $P(\sim e|\sim a.\sim c) = 1$. Thus, an observation of neither cause nor effect implies that it is less likely that the unobserved cause has happened. Again, for a definite prediction assumptions about the prior odds have to be made. However, as in the cases in which the observable cause is present, the prior odds are the same in all cases in which the observable cause is absent (see Equations 8 and 9). The prior odds of the unobserved cause given an observation of either $e.\sim c$ or $\sim e.\sim c$ are $P(a|\sim c)$ / $P(\sim a|\sim c)$. This implies that the posterior odds of the unobserved cause are lower when none of the events is observed than when the effect is observed by itself.

How can the prior odds be estimated? As we have shown in the first analysis the observable probability of the effect in the cause's absence provides a lower boundary for the probability of the unobserved cause, $P(a|\sim c) = 1 - P(\sim a|\sim c) \geq P(e|\sim c)$. Thus, after observing several trials at least the prior odds for cases in which the observable cause is absent can be estimated from the data.

The two analyses provided in this section show that even if no assumption of independence among the observed and the unobserved causes of a common effect is made, it is possible to derive some approximate inferences about the unobserved cause. This is true for individual trials as well as for complete data sets. Thus, informed guesses about unobserved causes can be made. For these analyses only probability calculus and Bayes' theorem were used. Consequently causal Bayes net theories can be adapted to cases in which the assumption of independence of causes seems unwarranted.

Summary

Current theories of causal learning consider unobserved causes and make assumptions that allow it to make inferences about these causes. Associative theories assume that an unobserved cause (i.e., the constant background) is always present and therefore independent of observable causes. Causal Bayes net theories including the special case of Power PC theory typically assume that different causes are independent of each other. This assumption allows these theories to make precise estimates about unobserved causes. However, despite the fact that these theories in general assume independence among different causes of a joint effect, these theories can also model situations in which the independence assumption does not hold. The analyses in the previous section have shown that in the case of dependence between observable and unobservable causes (i.e., confounding) it is normatively not possible to derive precise quantitative inferences about the probability and strength of the unobserved causes. Nevertheless it is possible to use Bayesian inferences to provide reasonable estimates. Even if no assumptions about dependence are made, the probability of the effect in the absence of the cause marks a lower boundary for the probability of the unobserved cause in the observable cause's absence, $P(a|\sim c) \geq P(e|\sim c)$.

Associative as well as causal Bayes net theories agree that $P(e|\sim c)$ is to a certain extent indicative of the causal strength of the unobserved cause. But whereas associative theories generally regard this probability as a valid indicator, Power PC and other causal Bayes net theories view this conditional probability as a lower boundary of the causal strength of the unobserved cause.

Causal Bayes net theories are also able to make specific predictions for different patterns of observations by using Bayesian inferences about unobserved causes (see previous section for details). If a cause and an effect are both observed it is more likely that an unobserved second cause is also present than when the cause is observed without an effect, $P(a|e.c) \geq P(a|\sim e.c)$. The existence and presence of an unobserved cause is certain if an effect is observed without any visible cause,

$P(a|e.\sim c) = 1$. If neither cause nor effect is observed, the presence of an unobserved cause is unlikely.

Experimental evidence

A number of researchers have provided evidence that people can and do infer the existence of unobserved causes. For example, children and adults infer the unobserved mental causes (e.g., desires and beliefs) of other people's actions (Gopnik & Wellman, 1994; Wellman, 1990). People also assume that there is an essential unobserved entity in categories, which has been speculated to play the role of a unobserved common cause of the observable features of category members (Gelman, 2003; Medin & Ortony, 1989; Rehder, 2003; Rehder & Hastie, 2001). In scientific reasoning people often infer unobserved entities to account for the observed events (Gopnik & Meltzoff, 1997; Keinath & Krems, 1999; Krems & Johnson, 1995). An example for recent studies on unobserved causes is a series of experiments conducted by Gopnik and colleagues (Gopnik et al., 2004; Kushnir et al., 2003). In one of their studies children were presented a stick-ball machine where they saw two balls moving together up and down for four times. Next an experimenter moved each ball by hand, which did not have any visible effect on the other ball. Children were asked to give an explanation. In accordance with causal Bayes net theories children as young as $4\frac{1}{2}$ years were able to reason that neither ball is the cause of the other so that there must have been an unobserved common cause that makes the two balls move together. These and other results show that children (and adults) are able to correctly infer the presence of an unobserved cause from a mixture of observations and interventions.

Our focus differs from these studies, however. Whereas Gopnik and colleagues (2004) were mainly interested in the induction of the presence of an unobserved common cause of multiple effects, our goal is to study inferences about the probability and the causal strength of unobserved causes that are known to exist. Moreover, whereas previous studies have focused on unobserved common causes of multiple observed effects, our experiments explore inferences about common-effect structures in which one cause is unobserved.

To the best of our knowledge only one study has thus far investigated some of these questions. Luhmann and Ahn (2003) have confronted participants with a single causal relation consisting of one observable cause and one observable effect (see also Ahn, Marsh, & Luhmann, in press). They showed that many participants were willing to judge the causal strength of unobserved causes that they had never observed during causal learning. In their first experiment Ahn and colleagues also found that the estimated causal influence of the unobserved causes declined when their number increased (the observable causal relation was kept constant). In a second study Luhmann and Ahn manipulated the probability of the observable effect when the observable cause was absent. It turned out that participants judged the causal strength of a single unobserved cause to be higher if $P(e|\sim c)$ was .5 than if it was zero. This finding is consistent with the theoretical accounts introduced above. Unfortunately, no estimates for the dependency between the observable and the unobserved cause were collected in these experiments. Therefore it is not possible to differentiate between an associative and a Bayesian account.

In an unpublished experiment (summarized in Ahn et al., in press) participants' inferences about the unobserved cause's presence were measured on each learning trial. It turned out that participants assumed the unobserved cause to be present when the effect had occurred on its own and to be absent when neither the visible cause nor the effect had occurred. This finding is in accordance with the predictions derived from a Bayesian analysis.

With the experiments presented in this paper we intended to go beyond these previous findings. In addition to causal strength estimates, we also collected assessments of the probability of the unobserved cause in the presence and absence of the observed cause and effect. We used a variety of methods to measure people's assumptions about the interdependence between observed and unobserved causes in different task contexts.

OUTLINE OF EXPERIMENTS

The aim of the following two experiments was to explore what assumptions participants make about the presence and causal strength of an unobserved cause in a trial-by-trial learning task and to test whether these assumptions conform to the predictions of any of the discussed theoretical models. In both experiments participants learned about the causal relation between an observable cause and an effect. In addition, participants were told that there was one other possible but unobserved cause of the effect. Participants never received any further information about the unobserved cause beyond this hint.

The statistical relation between the observable cause and the effect was manipulated in two experiments. In Experiment 1 the contingency between the observable cause and the effect was kept constant while the causal power varied across conditions. In Experiment 2 causal power was kept constant while the contingency varied. We used this manipulation to investigate whether participants would base their estimates of causal strength on the contingency between cause and effect or on the causal power of the cause, which could be inferred from the observations if independence amongst causes is assumed. Previous research yielded mixed evidence about this issue. While some researchers found support for the use of causal power (Buehner & Cheng, 1997; Buehner, Cheng, & Clifford, 2003) others found support for the use of contingencies (Lober & Shanks, 2000). Since we were interested in causal strength estimates, we were interested in finding out whether learners are sensitive to contingencies or causal power when estimating causal strength.

The manipulation of the statistical relation between the observable cause and the effect should also affect participants' guesses about the unobserved cause's strength. As we have outlined in the Introduction, the probability of the effect in the observable cause's absence is an indicator of the unobserved cause's strength and provides a lower boundary. Thus, the estimated causal strength should rise in parallel to the probability of the effect in the absence of the observed cause.

We also manipulated whether participants could only passively observe the cause and the effect (Experiment 1), or whether they could actively intervene and decide when the cause should be present or absent (Experiment 2). Mere observations without interventions do not permit it to make strong assumptions about the interdependence between the alternative causes. Either is possible, dependence or independence. Thus, participants may assume that the observed and the unobserved causes are dependent. In contrast, interventions should typically generate independence. Since the presence or absence of unobserved causes is unknown, deliberately setting a cause normally creates independence between the two causes (see Pearl, 2000; Spirtes et al., 1993; Woodward, 2003, for theoretical analyses of interventions). Therefore participants should assume independence (see Gopnik et al., 2004; Hagmayer, Sloman, Lagnado, & Waldmann, in press; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005, for psychological evidence).

We used two different tasks to assess participants' estimates of the probability of the unobserved cause: First, participants were asked to guess on each learning trial whether the unobserved cause was present or absent. Then, at the end of the learning session we requested additional summary estimates. Both measures were used to assess participants' assumptions about independence between observed and unobserved causes. By making several predictions in a series of trials participants generate patterns of alternative causes, which allow analysis of the dependence between the causes that participants implicitly generate. It is known in the literature on learning that participants tend to match the probability of events when generating sequences (see reviews by Hernstein, 1997; Myers, 1976; Vulkan, 2000). Therefore we hypothesized that participants' predictions would mirror their assumptions about the probability of the unobserved cause. Participants' final overall estimates are more explicit; they are based on summary information stored in memory. We were interested in whether these two types of measures would yield similar

estimates, or whether they would diverge. Since the online predictions are based on single cases, it may well be that these estimates focus on different aspects of the data from those focused on by the summary statements (see Catena, Maldonado, & Cándido, 1998, for an example of such a dissociation in the domain of causal strength estimation). For causal strength assessments summary estimates are, in our view, the more relevant evidence for independence assumptions as both are collected at the end of the learning phase.

EXPERIMENT 1

Participants' task in the first experiment was to learn about the causal relation between a microbe infecting flowering plants and the discoloration of their bloom. In addition, participants were requested to draw inferences about the presence and causal influence of a second but unobserved type of microbe, which may also affect the plant's coloration. This first experiment investigated observational learning—that is, participants could only observe whether a flower was infected and discoloured; they were not able to intervene themselves. Two experimental factors were manipulated: The first factor was the strength of the causal relation between the observable microbe and the discoloration of three types of plants. This factor was manipulated within subjects in a counterbalanced order. Contingencies between infection and discoloration were kept constant, but the two conditional probabilities $P(e|c)$ and $P(e|\sim c)$ were increased across conditions. This increase has two implications: (a) it implies an increasing causal power of the observable cause, and (b) it implies a higher probability or stronger causal strength of the second unobserved microbe because the probability of discoloration in the absence of an infection by the observable microbe increased.

As a second factor, which was manipulated between subjects, the learning procedure was manipulated: In two experimental conditions participants were requested to make predictions about the unobserved cause on each learning trial. In one of these conditions ("prediction after effect") participants received information about the presence of the first microbe and the effect and then had to predict the presence of the unobserved microbe. This information should allow them to make more informed guesses about the unobserved cause (see the analysis in the section on inferring unobserved causes). Summarizing our analyses in the Introduction, participants should be quite sure that the unobserved microbe is present if the observable microbe is absent but the flower is discoloured. They should also feel confident that the unobserved cause is absent if neither the observable cause nor the effect has occurred. If cause and effect were both present, learners should predict the unobserved cause to be present more often than if the cause was present but the effect was missing. No feedback was provided about the unobserved cause.

In the second experimental condition ("prediction before effect") participants had to make their guess after observing the presence or absence of the observable cause, but before being informed about the occurrence of the effect. Thus, they were first informed about the presence or absence of the cause in each trial, and then they were asked to guess whether the unobserved cause was present without receiving feedback about the alternative unobserved cause. Only after they had made their prediction were they informed whether the effect had occurred at this particular trial or not. In contrast to the "prediction after effect" condition this condition does not allow specific inferences to be made about the unobserved cause's presence on an individual trial. Participants' inferences of the unobserved cause prior to effect information can only be guesses based on observed frequencies of the effect in the absence of the observable cause in past trials. As we have outlined above, the probability of the effect in the absence of the cause provides a lower boundary of the base rate of the unobserved cause. Therefore participants should use the trials in which the observed cause is absent to infer the probability of the unobserved cause. Assuming independence they should adapt their predictions for the trials in which the observed cause is present to this

probability. If they did that, they would generate independence between the causes. A third control condition presented the learning data without requesting trial-by-trial predictions.

After observing the causal relation participants were requested to rate the causal strength of the observed as well as of the unobserved microbe. They were asked to estimate the probability of the second microbe being present, when the first microbe was present and when the first microbe was absent. Both participants' online inferences during learning and their final estimates were used to calculate their assumptions about the dependence between the two alternative causes.

Method

Participants and design
A total of 36 students from the University of Göttingen, Germany, participated in the experiment and were randomly assigned to one of the three learning conditions.

Materials and procedure
Participants were first instructed to learn about the causal relation between a fictitious microbe ("colorophages") and the discoloration of flowers. They were also told that there was only one other possible cause of the effect, an infection with another fictitious microbe ("mamococcus"), which was currently not observable. Next participants were presented a stack of index cards providing information about individual flowers. The front side of each index card showed whether the flower was infected by colorophages or not, and the backside informed about whether the flower was discoloured or not. Then participants were instructed about the specific learning procedure in their condition. In the "prediction before effect" condition participants were first shown the front side of the card, then they had to guess whether the flower was also infected by the other unobserved microbe, and finally the card was turned around by the experimenter revealing whether the flower was in fact discoloured or not. In contrast, in the second experimental

condition ("prediction after effect") the card was turned around after the presentation of the front side revealing the presence or absence of discoloration. Then the participants were asked to make their guesses about the unobserved cause. Guesses were secretly recorded without giving any feedback. In the third, the control, condition the cards were simply shown and turned around by the experimenter. No inferences about the unobserved cause were requested.

After each learning phase participants were asked to rate the strength of the causal influence of the observed and the unobserved cause on a scale ranging from 0 ("no causal influence") to 100 ("deterministic causal influence, i.e., the cause always yields the effect"). Participants were also asked to estimate how many of 10 flowers that were infected with the observed microbe were also infected with the other microbe, and how many of 10 flowers that were not infected with the observed microbe were instead infected with the other microbe. No feedback was provided about these assessments.

Three data sets were constructed in a manner that the contingency ΔP was held constant across all sets, whereas both $P(e|{\sim}c)$ and causal power were rising. Each data set consisted of 20 cases. The data sets and their statistical properties are shown in Table 1. All three data sets were shown to every participant in a within-subjects design. Different data sets were introduced as data from different species of flowers. It was pointed out to participants that the effectiveness of the microbes might vary depending on the species. The order of the presented data sets was counterbalanced.

Table 1. *Data sets shown in Experiment 1 and their statistical properties*

| | Data set | | | | Data set | | |
Observations	1	2	3	Statistics	1	2	3
$e.c$	6	8	10	$P(c)$	.50	.50	.50
${\sim}e.c$	4	2	0	$P(e|c)$	.60	.80	1.0
$e.{\sim}c$	1	3	5	$P(e|{\sim}c)$	.10	.30	.50
${\sim}e.{\sim}c$	9	7	5	ΔP	.50	.50	.50
Sum	20	20	20	Power p_c	.56	.71	1.0

Table 2. *Results of Experiment 1: Mean ratings of causal influence for observed and unobserved causes*

	Causal strength: Observed cause			Causal strength: Unobserved cause		
Condition	Data Set 1	Data Set 2	Data Set 3	Data Set 1	Data Set 2	Data Set 3
"Before effect"	57.5 (22.2)	68.3 (15.9)	78.3 (19.9)	50.5 (30.7)	48.3 (23.7)	46.7 (27.1)
"After effect"	64.1 (29.1)	70.8 (21.1)	77.9 (23.7)	34.2 (22.3)	61.7 (33.0)	73.8 (27.1)
Control	57.5 (19.1)	75.8 (12.4)	84.2 (16.2)	43.3 (25.6)	38.3 (21.7)	54.2 (19.3)

Note: Standard deviations in parentheses.

Results

First the estimated causal influence of the observed and the unobserved causes was analysed. Table 2 shows the mean ratings. Overall the descriptive means indicate increasing estimates for both causes. Separate analyses of variance for each cause with learning condition (before effect, after effect, control) as a between-subjects factor and data set (1, 2, 3) as a within-subjects factor were conducted. The analysis for the observed cause revealed a significant increase in causal strength ratings, $F(2, 66) = 12.7$, $MSE = 296.6$, $p < .01$, while the other main effect and the interaction failed to reach significance ($Fs < 1$). This pattern of results supports the predictions of Power PC theory, as the contingency between the cause and the effect remained constant across conditions. The analysis for the unobserved cause also resulted in a significant main effect of the factor data set, $F(2, 66) = 4.92$, $MSE = 408.1$, $p < .05$, which indicated that with increasing $P(e|\sim c)$ participants tended to assume a stronger causal strength of the unobserved cause. This result conforms to the predictions of all theoretical accounts discussed in the Introduction.

The interaction between data sets and learning condition also turned out to be significant,

$F(4, 66) = 4.55$, $MSE = 408.1$, $p < .05$. The observed increase was strongest in the "prediction after effect" condition followed by the control condition. This interaction can be explained by the different affordances of the three learning conditions. In the "prediction after effect" condition participants were sensitized to the possible presence and causal strength of the unobserved cause more than in the other two conditions. As we have shown in the Introduction, observing both cause and effect in a single trial allows informed predictions regarding the unobserved cause to be derived. No such predictions were possible in the other two conditions. In the "prediction before effect" condition participants lacked the information necessary to make an inference, which might have signalled to them that it is impossible to draw any inferences. In the control condition, they were never asked about the unobserved cause during learning.

Secondly, participants' implicit assumptions about the dependence between the causes were analysed. The trial-by-trial predictions of the unobserved cause in the presence and absence of the target cause were transformed into conditional frequencies and combined into subjective contingencies, $\Delta P = P(a|c) - P(a|\sim c)$.[1] This procedure

---

[1] We used the following formulae to derive $P_{gen}(a|c)$ and $P_{gen}(a|\sim c)$ from the observed frequencies with which different patterns of events were predicted:

$$P_{gen}(a|c) = [P_{gen}(a.e.c) + P_{gen}(a.\sim e.c)]/$$
$$[P_{gen}(a.e.c) + P_{gen}(\sim a.e.c) + P_{gen}(a.\sim e.c) + P_{gen}(\sim a.\sim e.c)]$$
$$P_{gen}(a|\sim c) = [P_{gen}(a.e.\sim c) + P_{gen}(a.\sim e.\sim c)]/$$
$$[P_{gen}(a.e.\sim c) + P_{gen}(\sim a.e.\sim c) + P_{gen}(a.\sim e.\sim c) + P_{gen}(\sim a.\sim e.\sim c)]$$

Table 3. *Results of Experiment 1: Mean generated and estimated dependencies of observed and unobserved causes*

	Generated dependence			Estimated dependence		
Condition	Data Set 1	Data Set 2	Data Set 3	Data Set 1	Data Set 2	Data Set 3
"Before effect"	.33 (.377)	.08 (.549)	−.22 (.273)	.30 (.252)	.06 (.360)	−.03 (.337)
"After effect"	.17 (.237)	.18 (.396)	−.29 (.204)	−.01 (.318)	.00 (.341)	−.23 (.470)
Control	—	—	—	.14 (.271)	.12 (.269)	.13 (.234)

Note: Values represent contingencies. Standard deviations in parentheses.

was based on the assumption that participants' predictions would reflect their underlying probability estimates due to probability matching. On the left side of Table 3 the mean contingencies that were generated online are listed. Participants' final ratings of the conditional frequency of the unobserved cause in the presence and absence of the observed cause were also transformed into subjective contingencies. Means are shown on the right hand side of Table 3. An analysis of variance of the generated contingencies with the between-subjects factor "learning procedure" and the within-subjects factor "data set" yielded a significant trend from positive to negative assessments, $F(2, 44) = 22.1$, $MSE = 749.8$, $p < .01$. No other main effect or interaction was found. Thus, the negative trend proved to be independent of the learning condition. The estimated contingencies showed a similar negative trend, $F(2, 66) = 4.2$, $MSE = 753.9$, $p < .05$. They also differed slightly across learning conditions, $F(2, 33) = 2.8$, $MSE = 1,646.4$, $p < .10$, but the interaction failed to reach significance.

To follow up these results more closely, generated and estimated contingencies were compared to an assumption of independence—that is, the empirically obtained contingencies were tested for deviations from zero using one-sample t tests. The analyses showed that the contingencies generated online were significantly above zero if $P(e|{\sim}c)$ was .1 (Data Set 1) and significantly below zero if $P(e|{\sim}c)$ was .5 (Data Set 3).

The pattern for the final summary estimates differed from the pattern for the generated dependencies. In comparison to an assumption of independence, only marginally significant deviations

were found for most of the estimates. The estimates for Data Set 1 in the "prediction before effect" condition were the only exception. These rather small deviations are consistent with an assumption of independence. This result points to a dissociation between online and post hoc assessments. While participants generated a positive dependence when the effect had never occurred in the absence of the observed cause and a negative dependence when it had occurred fairly often in the absence of the observable cause, the final summary estimates hovered around independence.

In a third analytical step the predictions regarding the unobserved cause, which participants made online on individual trials, were analysed. For this analysis the patterns of events generated in the online judgements were transformed into subjective probabilities of the unobserved cause conditional upon the patterns of the observed events. This means that they were transformed into probabilities of A conditional on both cause C and effect E, $P_{gen}(A|C, E)$, in the "prediction after effect" condition and into probabilities of A conditional on C, $P_{gen}(A|C)$, in the "prediction before effect" condition. Table 4 shows the mean conditional probabilities derived from the generated patterns. Note that in Data Set 3 participants never observed the effect to be absent when the observable cause was present. The probabilities in the "prediction after effect" condition conformed fairly well to the theoretical predictions derived in the Introduction. Participants judged the unobserved cause to be much more likely when the effect had occurred in the observable cause's absence than when neither cause nor effect had

Table 4. *Results of Experiment 1: Mean probabilities of unobserved cause generated by participants conditional upon each pattern of observations*

"Before effect" condition											
Data Set 1				Data Set 2				Data Set 3			
$P_{gen}(a\|c)$		$P_{gen}(a\|\sim c)$		$P_{gen}(a\|c)$		$P_{gen}(a\|\sim c)$		$P_{gen}(a\|c)$		$P_{gen}(a\|\sim c)$	
.66 (.22)		.32 (.20)		.50 (.35)		.53 (.26)		.27 (.16)		.48 (.21)	
"After effect" condition											
Data Set 1				Data Set 2				Data Set 3			
$P_{gen}(a\|e.c)$	$P_{gen}(a\|\sim e.c)$	$P_{gen}(a\|e.\sim c)$	$P_{gen}(a\|\sim e.\sim c)$	$P_{gen}(a\|e.c)$	$P_{gen}(a\|\sim e.c)$	$P_{gen}(a\|e.\sim c)$	$P_{gen}(a\|\sim e.\sim c)$	$P_{gen}(a\|e.c)$	$P_{gen}(a\|\sim e.c)$	$P_{gen}(a\|e.\sim c)$	$P_{gen}(a\|\sim e.\sim c)$
.40 (.34)	.21 (.35)	.92 (.29)	.07 (.12)	.52 (.43)	.29 (.40)	.92 (.15)	.05 (.16)	.44 (.38)	—	1.0 (0.0)	.02 (.06)

Note: Standard deviations in parentheses.

occurred. In fact, participants in this condition were almost certain that the unobserved cause was present, when the observable cause was absent, $P_{gen}(a|\sim c.e) > .90$, and absent otherwise, $P_{gen}(a|\sim c.\sim e) < .10$. An analysis of variance with the two within-subjects factors "data set" (1, 2, 3) and "presence of effect" (present vs. absent) yielded a highly significant main effect of the factor involving the effect's presence, $F(1, 11) = 347.0$, $MSE = 0.042$, $p < .01$. No other effect turned out to be significant. Participants' predictions were less clear cut when the observable cause was present. At least the descriptive differences pointed in the right direction. Participants considered the unobserved cause to be more likely when the observable cause and the effect had occurred than when only the observable cause had occurred without the effect. However, an analysis of variance with the two within-subjects factors "data set" (1 vs. 2) and "presence of effect" (present vs. absent) yielded no significant effect of the factor "presence of effect", $F(1, 11) = 2.59$, $MSE = 0.208$, $p = .14$. Increasing the statistical power of the experiment might confirm the expected difference (see Experiment 2). Nevertheless, participants' inferences exhibit a surprising grasp of the inferences that can be drawn from the individual patterns of events.

A rather awkward pattern resulted in the "prediction before effect" condition. Contrary to our predictions participants did not predict the unobserved cause to be present more often, when the probability of the effect in the observable cause's absence rose. They rather predicted the unobserved cause to be present less often in the observables cause's presence when the causal impact of the observable cause increased. In contrast to the "prediction after effect" condition it seems that participants had no clue how to derive predictions about the unobserved cause in this condition.

Overall the results from Experiment 1 contradicted some of the theoretical assumptions made by the currently predominant theories of causal learning. Participants assumed neither that the unobserved cause was always present nor that the two causes were independent (at least in the online judgements). This finding challenges

associative theories, Power PC theory, and other causal Bayes net theories, which assume independence of causes. However, we already have pointed out that an independence assumption is not necessary for Power PC theory and causal Bayes net theories in general. Causal Bayes net theories including Power PC theory can model dependence between observed and unobserved causes. Therefore, even if participants' answers did not conform to the independence assumption, their answers still might be coherent with modified versions of these theories. To be coherent participants' estimates would have to honour the constraints imposed by the learning data. The most important constraint is that the product of the causal power (or causal strength) of the unobserved cause and the probability of the unobserved cause in the absence of the observed cause must equal the probability of the effect in the absence of the observed cause.

To find out whether participants honour this constraint we used their causal strength and their final summary ratings of the unobserved cause to recalculate the probability of the effect E when cause C was absent:

$$P_{rec}(e|\sim c)$$
$$= \text{causal strength rating} A^* \text{rating } P(a|\sim c).$$

The recalculated probabilities and the actually observed probabilities are shown in Table 5. It can be seen that the recalculated probabilities in the "prediction after effect" condition were surprisingly close to the actually observed probabilities. In contrast, the recalculated probabilities in the other two conditions were inaccurate. Since these ratings were collected at the end of the learning phase the necessary information to make accurate summary estimates was available to learners in all conditions. However, it seems that participants had to be sensitized by the learning procedure to the presence and causal strength of the unobserved cause to be able to derive coherent estimates. The procedure in the "prediction after effect" condition required participants in each trial to think about the unobserved cause, and it provided participants with the right

Table 5. *Recalculated conditional probabilities of the effect in the absence of the observable cause*

Condition	Data Set 1		Data Set 2		Data Set 3	
	Recalculated	Observed	Recalculated	Observed	Recalculated	Observed
"Before effect"	.13 (.13)		.25 (.19)		.24 (.17)	
"After effect"	.14 (.15)	.10	.26 (.26)	.30	.43 (.28)	.50
Control	.14 (.10)		.21 (.20)		.21 (.14)	

Note: Calculations are based on the probabilities and causal strength parameters estimated by participants. Power PC theory is used to integrate these estimates. See text for more details. Standard deviations in parentheses.

information to be able to do so. Without this information participants' guesses showed some systematicity but they did not conform very well to the observed data. Thus, simply asking participants to think about unobserved causes on each trial—as we did in the "prediction before effect" condition—was clearly not sufficient.

An interesting novel finding is the dissociation between online judgements and final summary judgements. Although the online judgements show deviations from the independence assumption, the final judgements are consistent with this assumption. Apparently, online judgements were driven by other aspects of the data than the summary estimates. We discuss this point more closely in the General Discussion. This finding supports normative theories of power estimation (e.g., Power PC theory) that hypothesize that people default to an independence assumption. These theories might argue that the online judgements are less relevant than the final estimates, as power estimates are typically not made online but rather on the basis of a learning sample. If people retrospectively make the independence assumption when assessing power, their estimates should correspond to the predictions of normative theories (i.e., Power PC theory). This point is reinforced by the fact that learners' estimates seemed to mirror causal power rather than contingencies.

EXPERIMENT 2

In Experiment 1 participants could only passively observe the occurrence of a cause and an effect.

Mere observations do not rule out the possibility that the observed cause is in fact related to a second unobserved, confounding cause. It might have been the case that flowers were more often infected by both microbes than by only one. Such dependence might have even been plausible for some participants. To make independence more salient, we switched from observations to interventions in Experiment 2. In this experiment we allowed participants to arbitrarily manipulate the observable cause. Since these random interventions cannot be based on the presence or absence of the unobserved cause, they are more likely to be independent from the unobserved cause (see Hagmayer et al., in press; Meder, Hagmayer, & Waldmann, 2005; Waldmann & Hagmayer, 2005). Therefore, learning from interventions should make independence between alternative causes more salient than should learning from observations. We speculated that learners may be more prone to assume independence between alternative causes if they are allowed to freely manipulate the observable cause.

Participants in this second experiment were instructed to imagine being a captain on a pirate ship firing his battery at a fortress. Their task was to assess their own hit rate—that is, the causal strength of their actions upon the fortress. A second ship, which also fired at the fortress but was occluded from participants' view, served as the unobserved cause. Participants had to decide whether to fire or not on each trial. Only a limited number of shells were provided to ensure that all participants received equivalent data despite the fact that they arbitrarily set the cause themselves.

The same two factors as those in the first experiment were manipulated. Learning conditions were again varied between subjects. In the two experimental conditions participants either had to guess whether the other ship had fired during the current trial before they were informed about the occurrence of an explosion in the fortress ("prediction before effect"), or they had to predict the other ship's action after they had learned whether the fortress was hit ("prediction after effect"). In a third control condition no predictions were requested. At the end of the learning phase all participants were requested to give explicit estimates of the probability of the unobserved cause conditional upon the manipulated cause. They also had to rate the causal strength of the observed and of the unobserved cause.

As a second factor the data sets presented to participants were varied as a within-subjects variable. The order of the data sets was counterbalanced as in the first experiment. Again the conditional probability of the effect in the presence and absence of the cause was raised across conditions. In contrast to Experiment 1, in which contingencies were kept constant, we now kept causal power constant. Therefore we expected participants to rate the causal strength of the manipulated cause to be the same in all conditions. However, we again expected that the estimated probability and causal strength of the unobserved cause would rise in parallel to the probability of the effect in the absence of the observable cause.

Method

Participants and design
A total of 60 students from the University of Göttingen, Germany, participated and were randomly assigned to one of the three learning conditions.

Materials and procedure
This second experiment was run on a computer. First, participants were instructed to imagine being a captain on a pirate ship trying to invade fortresses in the Caribbean. Therefore they were firing at a fortress with their ship's battery. A second friendly ship, not visible from the captain's current position, allegedly also fired at the fortress. Participants were told that they had the chance to fire 30 salvos at the fortress on 60 occasions. Their task was to assess how often their own battery would hit the fortress. On each trial participants were first asked whether they wanted to fire a salvo until they used up their shells. In the "prediction before effect" condition they were next asked to predict whether the other ship had fired on this occasion or not. No feedback was provided. Then learners observed whether a causal effect (a blast within the fortress) had occurred or not. In the "prediction after effect" condition participants were informed after their intervention whether an explosion at the fortress had occurred or not, and then had to predict whether the other ship had also fired on this occasion. Again no feedback was given. In the control condition participants only decided whether to fire and then observed the causal effect.

After completing the 60 trials participants were asked to rate how often their salvos would hit the fortress if the other ship had stopped firing. Participants gave their answer on a scale ranging from 0 (= "never") to 100 (= "always"). They also had to estimate how often the other ship's battery would hit the fortress if they had stopped firing themselves. The same rating scale was used again. Participants were also asked how many of the 30 times they had fired had the other ship fired as well and how often within the set of 30 trials in which they had not fired had the other ship fired instead. No feedback was provided. Participants were then told that they had successfully captured the fortress and sailed on to the next. Before the next data set was shown, it was emphasized that the environmental conditions at the new fortress were completely different, so that other success rates might result.

Three new data sets consisting of 60 cases each were constructed. Each was constructed in a way that the contingency between the observed cause and the effect decreased across the data sets, whereas the causal power remained constant. Table 6 shows the conditional probabilities that were used to generate the data for each participant

Table 6. *Conditional probabilities used to generate the three data sets of Experiment 2*

	Data Set 1	Data Set 2	Data Set 3
$P(e|c)$	.70	.80	.90
$P(e|\sim c)$	.00	.33	.67
ΔP	.70	.47	.23
Power p_c	.70	.70	.70

in the three data set conditions. Each set was generated individually for each participant. This means that the data sets were not fixed in advance like in Experiment 1 but that the presence of the effect was determined on each trial anew using the probabilities shown in Table 6. For example, if a participant in the Data Set 1 condition chose to fire, the computer reported a causal effect with a probability of $P = .7$. As in Experiment 1 the order of the data sets was counterbalanced across participants.

Results

First participants' estimates of causal strength of the manipulated and the unobserved cause were analysed. Mean ratings for all conditions are shown in Table 7. An analysis of variance of the causal strength ratings of the manipulated cause with learning condition as a between-subjects factor and data sets as a within-subjects factor yielded no significant effects. This result is in accordance with the predictions of Power PC theory because causal power was kept constant across data sets. The same analysis conducted for the causal strength estimates concerning the unobserved cause yielded two significant main effects.

As in Experiment 1 the estimated causal influence rose across the three data sets, $F(2, 114) = 65.7$, $MSE = 408.2$, $p < .01$. This finding is consistent with all theories considered in the Introduction. There was also a significant difference between learning conditions, $F(2, 57) = 4.06$, $MSE = 591.8$, $p < .05$. Participants in the "prediction after effect" condition rated the causal strength of the unobserved cause to be higher than that in the other two conditions. This result may be a consequence of the learning procedure in this condition, which drew participants' attention to the possible influence of the unobserved cause. Remember that participants in this condition had to predict the other cause's presence based on their own action and the occurrence of the effect. Therefore every time the effect had occurred without the participant's intervention, the participant should have concluded that the other cause must have been present. No such predictions were required in the other two conditions, which might be the reason why participants in these conditions may have overlooked at least some of the crucial cases in which the effect had occurred on its own.

In a second step participants' implicit assumptions about the dependence between the manipulated and the unobserved cause were analysed. Thus again, the conditional frequencies generated by participants during their trial-by-trial predictions were translated into conditional probabilities and subtracted to yield contingencies. The resulting subjective dependencies are listed in Table 8. Supporting the results of Experiment 1, participants tended to generate a negative dependence between the causes. With one exception (Data Set 1 in the "prediction after effect" condition) all

Table 7. *Results of Experiment 2: Mean ratings of causal influence of observed and unobserved causes*

Condition	Causal strength: Observed cause			Causal strength: Unobserved cause		
	Data Set 1	Data Set 2	Data Set 3	Data Set 1	Data Set 2	Data Set 3
"Before effect"	67.7 (22.7)	65.1 (20.9)	68.0 (18.6)	27.3 (23.3)	44.8 (20.6)	65.8 (20.3)
"After effect"	70.0 (29.1)	61.4 (17.2)	62.0 (32.5)	36.0 (26.6)	57.9 (17.0)	81.6 (13.1)
Control	66.9 (15.8)	63.6 (11.8)	71.2 (17.6)	33.9 (32.7)	50.6 (21.2)	76.3 (16.8)

Note: Standard deviations in parentheses.

Table 8. *Results of Experiment 2: Mean generated and estimated dependencies of observed and unobserved causes*

	Generated dependence			Estimated dependence		
Condition	Data Set 1	Data Set 2	Data Set 3	Data Set 1	Data Set 2	Data Set 3
"Before effect"	−.24 (.319)	−.33 (.380)	−.29 (.438)	−.04 (.203)	.08 (.256)	−.08 (.306)
"After effect"	−.03 (.300)	−.20 (.276)	−.35 (.258)	−.13 (.341)	−.11 (.304)	−.01 (.589)
Control	—	—	—	.10 (.366)	.09 (.314)	−.06 (.339)

Note: Values represent contingencies. Standard deviations in parentheses.

generated dependencies deviated significantly from zero. An analysis of variance with data set and learning conditions as factors yielded a significant main effect of data set, $F(2, 76) = 6.97$, $MSE = 510.1$, $p < .01$ and a significant interaction, $F(2, 76) = 3.57$, $MSE = 510.1$, $p < .05$. The dependence became more negative when participants had received effect information before their predictions but remained at the same negative level when they made their predictions before being informed about the effect. In contrast to the generated dependencies in the online judgements the estimated dependencies in the final summary judgements did not statistically differ from each other and from zero. Thus, as in the first experiment, there was a clear dissociation between online and final summary judgements. Whereas the online judgements showed deviations from the independence assumption, the final estimates corresponded to it.

In a third analytical step we again analysed the trial-by-trial online predictions more closely. As in Experiment 1, the generated patterns were transformed into subjective probabilities of the unobserved cause conditional upon the manipulated cause, $P_{gen}(A|C)$, in the "prediction before effect" condition and probabilities conditional upon patterns of the observed cause and the effect, $P_{gen}(A|C,E)$, in the "prediction after effect" condition. The results are shown in Table 9. As in Experiment 1 it is important to note that Data Set 1 contained no cases in which the effect occurred in the absence of the manipulated cause. Overall the results were similar to those in Experiment 1. The probabilities generated in the "prediction after effect" condition again conformed

to the implications of a Bayesian analysis. Participants predicted the unobserved cause to be present with a very high probability when the observable cause was absent in the effect's presence and predicted it with a rather low probability when both cause and effect were absent. An analysis of variance with the two within-subject factors "data set" (2 vs. 3) and "presence of effect" (present vs. absent) yielded a strong main effect of the presence of the effect, $F(1, 19) = 115.4$, $MSE = 0.071$, $p < .01$, while all other effects were insignificant. Participants also differentiated between the cases in which both cause and effect were present and the cases in which the cause was present without generating the effect. In accordance with the Bayesian predictions participants inferred the unobserved cause with a higher probability in the first than in the second case. An analysis of variance with the two within-subjects factors "data set" and "presence of effect" confirmed the significance of this difference, $F(1, 19) = 7.50$, $MSE = 0.111$, $p < .05$. The probabilities generated by participants in the "prediction before effect" condition show a uniform pattern. Learners predicted the unobserved cause to be present more often when the observed cause was absent than when it was present. There was apparently no difference between the three data sets. Recall that in this condition participants had to make their prediction before receiving information about the effect, which did not allow them to draw specific inferences for individual cases. Nevertheless, participants should have adapted their predictions to the observable probability of the effect in the observable cause's presence, which defines a lower boundary for the probability of the unobserved cause. This conditional

Table 9. *Results of Experiment 2: Mean probabilities of unobserved cause generated by participants conditional upon each pattern of observations*

		"Before effect" condition			
Data Set 1		Data Set 2		Data Set 3	
$P_{gen}(a\|c)$	$P_{gen}(a\|\sim c)$	$P_{gen}(a\|c)$	$P_{gen}(a\|\sim c)$	$P_{gen}(a\|c)$	$P_{gen}(a\|\sim c)$
.37 (.24)	.61 (.26)	.36 (.24)	.69 (.24)	.40 (.24)	.69 (.26)

				"After effect" condition							
Data Set 1				Data Set 2				Data Set 3			
$P_{gen}(a\|e.c)$	$P_{gen}(a\|\sim e.c)$	$P_{gen}(a\|e.\sim c)$	$P_{gen}(a\|\sim e.\sim c)$	$P_{gen}(a\|e.c)$	$P_{gen}(a\|\sim e.c)$	$P_{gen}(a\|e.\sim c)$	$P_{gen}(a\|\sim e.\sim c)$	$P_{gen}(a\|e.c)$	$P_{gen}(a\|\sim e.c)$	$P_{gen}(a\|e.\sim c)$	$P_{gen}(a\|\sim e.\sim c)$
.39 (.22)	.29 (.21)	—	.40 (.29)	.42 (.21)	.26 (.30)	.99 (.04)	.39 (.32)	.41 (.24)	.18 (.25)	.98 (.05)	.29 (.29)

Note: Results concerning the "before effect" condition are shown in the upper half of the table, results concerning the "after effect" condition are shown in the lower half. Standard deviations in parentheses.

Table 10. *Recalculated conditional probabilities of the effect in the absence of the observable cause*

Condition	Data Set 1		Data Set 2		Data Set 3	
	Recalculated	Observed	Recalculated	Observed	Recalculated	Observed
"Before effect"	.13 (.13)		.20 (.12)		.42 (.22)	
"After effect"	.17 (.19)	.00	.35 (.17)	.33	.55 (.21)	.67
Control	.10 (.15)		.24 (.13)		.55 (.18)	

Note: Calculations are based on the probabilities and causal strength parameters estimated by participants. Power PC theory is used to integrate these estimates. Standard deviations in parentheses.

probability increased across trials. Therefore we expected to find a difference between the data sets at least for trials in which the observable cause was absent. As in Experiment 1 participant's predictions showed no sensitivity to this implication.

Finally, we again investigated whether participants' final ratings honoured the constraints imposed by causal Bayes net theories and Power PC theory. The estimated probabilities and causal strength estimates were used to recalculate the probability of the effect in the absence of the observable cause. If participants were sensitive to the constraints imposed by these theories, the values should closely resemble the observed conditional probabilities. The results are shown in Table 10. It can be seen that participants tended to honour the constraints. As in Experiment 1 the best performance was found in the "prediction after effect" condition in which participants (a) were asked about the unobserved cause on each occasion, and (b) had the information available that allowed them to make informed guesses.

GENERAL DISCUSSION

All current theories of causal reasoning consider unknown causes. However, there is no agreement about the correct way to model inferences about such causes. The aim of this paper was to empirically investigate the assumptions that learners make about unobserved causes to provide constraints for theories of causal learning. In two experiments we requested participants to learn a single causal relation between an observable

cause and an effect that, according to our instructions, could also be caused by an unobserved cause. Participants never received any feedback about this unobserved cause. Nevertheless, we asked them to assess the relation between the observed and the unobserved cause and the causal strength of both the observed and the unobserved cause upon the effect.

Associative theories predict that the estimated causal influence of the observed cause should correspond to the observed contingency between the observed cause and the effect. Power PC theory and other causal Bayes net accounts, on the other hand, predict that the estimated causal influence of a cause should correspond to its causal power—that is, its capacity to produce the effect in the absence of all other potential causes. The results of both experiments supported Power PC theory over an associative account. Participants' estimates conformed to the pattern predicted by causal power (see also Buehner et al., 2003).

All theoretical accounts agree that the causal strength of the unobserved cause has to be at least as high as the probability of the effect in the observable cause's absence, $P(e|\sim c)$. However, whereas associative theories would predict that the estimated causal influence equals this probability, Power PC theory and other causal Bayes net accounts predict that this probability just marks the lower boundary of the admissible values of causal strength. We found evidence in both experiments that the estimates of causal strength of the unobserved cause increased proportional to $P(e|\sim c)$. This finding replicates previous results (Luhmann & Ahn, 2003) and is in

accordance with all theories that were discussed in the Introduction. However, we also found that the estimates were consistently higher than $P(e|\sim c)$. This result favours causal Bayes net theories over an associative account.

Associative theories assume that an unobserved cause is permanently present. This assumption ensures that the unobserved cause is independent of the observed cause, which allows the causal strength of both causes to be estimated. Causal Bayes net theories including Power PC theory are more flexible and allow the analysis of situations in which the unobserved cause can be present or absent. If independence is assumed, the causal power of observable causes can be precisely determined even when no information about the unobserved cause is available (see the power formula). In this case the probability of the unobserved cause and its causal strength are both constrained by the observable conditional probabilities, especially the probability of the effect in the observable cause's absence, $P(e|\sim c)$, which marks the lower boundary. In none of the experiments was evidence found that participants assumed that the unobserved cause was permanently present.

However, the results also did not unanimously support the independence assumption usually made by causal Bayes net accounts. In both experiments we found an interesting dissociation between participants' online and final summary judgements of the causes' interdependence. While their final estimates about the relation between the observed and the unobserved cause were consistent with an assumption of independence, their predictions for individual trials were not. In Experiment 1 (observational learning) participants generated a positive dependence if the effect only rarely occurred on its own, but a negative dependence if it occurred fairly often without any apparent cause. In Experiment 2 (interventional learning) participants generated a negative dependency throughout. This implies that participants thought that the unobserved cause had occurred more often when the observable cause was absent than when it was present. This finding is remarkable because interventions should ensure that the manipulated cause occurs independently of all unobserved causes.

There are several possible explanations for this unexpected finding of a negative dependency in online judgements. One reviewer speculated that the observed dependency might be due to the characteristics of the task. Predictions for individual trials require translating subjective probabilities into binary judgements. Such a translation process might generate a bias towards dependency. While we cannot rule out this possibility it seems unlikely to us. Our analyses are based on findings of an extensive literature that people tend to match probabilities when generating sequences of events (see Hernstein, 1997; Myers, 1976; Vulkan, 2000, for overviews). This assumption is supported by the fact that the predictions that were based on observed patterns of cause and effect ("prediction after effect" condition) closely mirrored the predictions derived from a Bayesian analysis. There was no sign of a bias in these judgements.

It is important to recall that the results concerning the online predictions clearly differed between the two learning conditions. The predictions made before information about the presence of the effect was revealed ("prediction before effect" condition) hardly showed any systematic relation to the data sets. The tendency to predict the unobserved cause less often when the observable cause was already present may be grounded in people's reluctance to accept overdetermination of an effect. Since one cause suffices to explain an effect, assuming a second unobserved cause is not necessary. A second related intuition is that events which are not causally related rarely occur simultaneously by chance. Thus, it seems unlikely that the observed and the unobserved cause co-occur. This would also explain an overall tendency to generate a negative dependency in this condition.

The tendency to create an increasingly negative dependency in the "prediction after effect" condition showed that participants in this condition were sensitive to the statistical properties of the data sets. In both experiments participants almost every time predicted the presence of the unobserved cause when the effect had occurred on its own. In addition, participants tended to assume its absence when neither the observable cause nor the effect had occurred. When the

observed cause and the effect were present participants ascribed a slightly higher probability to the unobserved cause than when the observed cause had occurred without the effect. However, the overall probability assigned to the unobserved cause was medium to low in both cases and varied only slightly across conditions. These findings for the individual patterns of observations explain why participants tended to generate an increasingly negative dependency in the "prediction after effect" condition across data sets. Across the data sets the number of trials in which the effect occurred on its own increased. Since participants assumed that some cause must be responsible for the observed effect, the rising number of these cases led to more predictions of the unobserved cause across data sets—that is, the probability of the unobserved cause in the observed cause's absence, $P_{gen}(a|\sim c)$, rose substantially. In addition, the number of trials in which cause and effect co-occurred also increased across data sets. However, this trend only slightly increased the frequency of predictions of the unobserved cause in the observed cause's presence because participants ascribed only a slightly higher probability to the unobserved cause in the presence of both cause and effect than in trials when the cause was present by itself (see above). Therefore $P_{gen}(a|c)$ rose only slightly. As a consequence a negative trend was observed across data sets. Thus, the finding of an increasingly negative dependency might be just a sign of participants' Bayesian reasoning about individual trials that neglects the statistical properties of the whole sequence of events.

While participants' online judgements deviated significantly from independence, their final summary estimates did not. This dissociation provides an interesting challenge for theories on causal learning (see also Catena et al., 1998, for a similar dissociation regarding causal strength estimates). As we have outlined in the previous paragraphs the dependence generated in the online judgements is probably due to the way the predictions were derived for each individual case. Summary estimates, however, most likely do not focus on individual cases but on larger samples of cases. More research is needed to explore the processes underlying this interesting dissociation.

What are the implications of these findings for causal Bayes net theories and Power PC theory? Since these theories model strength estimates obtained at the asymptote of learning, summary judgements at the end of learning may be the more valid indicator of participants' assumptions about unobserved causes at this point of learning. These estimates did not on average deviate from independence, which is consistent with Power PC theory.

Even if the online judgements were viewed as the more valid indicator of people's intuitive assumption about dependence, causal Bayes net theories or Power PC theory are not refuted. Both theories may drop this assumption, however, at the cost of making causal power no longer precisely estimable. Causal Bayes net theories can model cases in which unobserved and observed causes are dependent. In this case, these theories provide constraints for consistent estimates. We used participants' final estimates to find out whether the estimates honour these constraints. The most important constraint is that the probability of the effect in the absence of the unobserved cause equals the product of the probability of the unobserved cause in the absence of the observed cause and the causal power of the unobserved cause, $P(e|\sim c) = P(a|\sim c)p_a$ (see Equations 2 and 5). Participants' estimates in fact tended to honour this constraint. Thus, even if independence was not assumed by learners, their estimates were coherent with a rational Bayesian analysis.

What are the implications of these findings for the debate between associative and cognitive theories of causal learning? Our findings clearly challenge associative theories (see Shanks, 2007) because participants proved capable of drawing systematic inferences about the presence and causal strength of cues that remained unobservable throughout learning. This finding supports the assumption of rational theories of causal learning (e.g., Power PC theory) that people represent causal learning tasks as situations in which unobservable causes and observed causes jointly

generate the observed effects within a common-effect model. By contrast, associative theories claim that observed causes compete with the constantly present and observable context without separately considering unobservable causal events. Our results add to other findings reported in this volume, which show that participants use the observable information to draw rational inferences about observable and unobservable causal events and are sensitive to the structure of causal models (see Booth & Buehner, 2007; Cobos, López, & Luque, 2007; De Houwer, Vandorpe, & Beckers, 2007; Vandorpe, De Houwer, & Beckers, 2007).

To sum up, our results contradict the assumption of associative theories that learners assume the constant presence of alternative, unobserved causes. The results about the independence assumption made by Bayesian theories are mixed. Whereas the online estimates violated independence, occasionally in the direction of a positive, more often in the direction of a negative correlation, the summary estimates were on average close to independence. Moreover, a Bayesian analysis revealed that the estimates were rational and were consistent with Bayesian constraints. These analyses provide further convincing evidence for the usefulness of a Bayesian analysis of causal learning.

REFERENCES

Ahn, W.-K., Marsh, J. K., & Luhmann, C. C. (in press). Dynamic interpretations of covariation data. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford, UK: Oxford University Press.

Audi, R. (1995). *The Cambridge dictionary of philosophy*. Cambridge, MA: Cambridge University Press.

Booth, S. L., & Buehner, M. J. (2007). Asymmetries in cue competition in forward and backward blocking designs: Further evidence for causal model theory. *Quarterly Journal of Experimental Psychology, 60*, 387–399.

Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The power PC theory versus the Rescorla−Wagner model. In P. Langley & M. G. Shafto (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 55–60). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariaton to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1119–1140.

Catena, A., Maldonado, A., & Cándido, A. (1998). The effect of frequency of judgement and the type of trials on covariation learning. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 481–495.

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

Cobos, P. L., López, F. J., & Luque, D. (2007). Interference between cues of the same outcome depends on the causal interpretation of the events. *Quarterly Journal of Experimental Psychology, 60*, 369–386.

Danks, D. (2003). Equilibria of the Rescorla−Wagner model. *Journal of Mathematical Psychology, 47*, 109–121.

Danks, D. (in press). Causal learning from observations and manipulations. In M. Lovett & P. Shah (Eds.), *Thinking with data*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

De Houwer, J., Vandorpe, S., & Beckers, T. (2007). Statistical contingency has a different impact on preparation judgements than on causal judgements. *Quarterly Journal of Experimental Psychology, 60*, 418–432.

Gelman, S. (2003). *The essential child. Origins of essentialism in everyday thought*. Cambridge, MA: Oxford University Press.

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

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., & Meltzoff, A. N. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 285–386.

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, UK: Oxford University Press.

Hernstein, R. J. (1997). *The matching law*. Cambridge, MA: Harvard University Press.

Jensen, F. V. (1997). *An introduction to Bayesian networks*. London: Springer.

Keinath, A., & Krems, J. F. (1999). Anomalous data integration in diagnostic reasoning. In A. Bagnara (Ed.), *Proceedings of the 1999 European Conference on Cognitive Science* (pp. 213–218). Siena, Italy: Consiglo Nationale delle Ricerche.

Krems, J. F., & Johnson, T. (1995). Integration of anomalous data in multi causal explanations. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 277–282). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kushnir, T., Gopnik, A., Schulz, L., & Danks, D. (2003). Inferring hidden causes. In P. Langley & M. G. Shafto (Eds.), *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107*, 195–212.

Luhmann, C. C., & Ahn, W.-K. (2003). Evaluating the causal role of unobserved variables. In P. Langley & M. G. Shafto (Eds.), *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Meder, B., Hagmayer, Y., & Waldmann, Y. (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*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge, UK: Cambridge University Press.

Myers, J. L. (1976). Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Approaches to human learning and motivation* (pp. 171–205). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

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

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130*, 323–360.

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

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

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60*, 291–309.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (1996). *The psychology of learning and motivation: Vol. 34. Causal learning*. San Diego, CA: Academic Press.

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

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

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 15, pp. 35–42). Cambridge, MA: MIT Press.

Vandorpe, S., De Houwer, J., & Beckers, T. (2007). Outcome maximality and additivity training also influence cue competition in causal learning when learning involves many cues and events. *Quarterly Journal of Experimental Psychology, 60*, 356–368.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14*, 101–118.

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

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

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response–outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 174–188.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 710–727.

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