Psychological Review
2007, Vol. 114, No. 3, 657–677

Copyright 2007 by the American Psychological Association
0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.3.657

BUCKLE: A Model of Unobserved Cause Learning

Christian C. Luhmann and Woo-kyoung Ahn
Yale University

Dealing with alternative causes is necessary to avoid making inaccurate causal inferences from covariation data. However, information about alternative causes is frequently unavailable, rendering them unobserved. The current article reviews the way in which current learning models deal, or could deal, with unobserved causes. A new model of causal learning, BUCKLE (bidirectional unobserved cause learning) extends existing models of causal learning by dynamically inferring information about unobserved, alternative causes. During the course of causal learning, BUCKLE continually computes the probability that an unobserved cause is present during a given observation and then uses the results of these inferences to learn the causal strengths of the unobserved as well as observed causes. The current results demonstrate that BUCKLE provides a better explanation of people's causal learning than the existing models.

Keywords: causal learning, unobserved causes, missing data, Bayesian inference

In this article, we explore how people learn about causes when given incomplete information. Specifically, we discuss situations in which causes are unobserved. Many existing models of causal induction ignore unobserved causes (e.g., Anderson & Sheu, 1995; Busemeyer, 1991; Cheng & Novick, 1992; Jenkins & Ward, 1965; Schustack & Sternberg, 1981; White, 2002) or make simplistic assumptions about unobserved causes (e.g., Rescorla & Wagner, 1972). In contrast, we argue that learners use a more sophisticated strategy for dealing with unobserved causes. We present a new model, BUCKLE (bidirectional unobserved cause learning), to formalize our theory. Before presenting this model, we first discuss the need to deal with unobserved causes and previous attempts to do so.

The Importance of Alternative Causes

Causal beliefs are generally assumed to result from experience in the form of covariation: how the causes vary with their effects. The covariation between a single cause and effect can be summarized in a table like the one in Figure 1. Thus, a learner observes whether presence or absence of a (possible) cause is followed by presence or absence of the (possible) effect and translates these observations into beliefs about the intervening causal relationship. Much work has been dedicated to exploring how this translation is made (see, e.g., Shanks, Holyoak, & Medin, 1996, for an extensive review).

Work over the last 25 years has revealed that the covariation-to-causality translation is more complex than traditional theories suggested. In particular, it seems clear that beliefs about one cause critically depend on how learners deal with other, alternative causes of that same effect (e.g., Cheng, Park, Yarlas, & Holyoak, 1996). For example, Spellman (1996) had participants learn about two liquids (one red and one blue) and their influence on flowers' blooming. When participants were asked about the influence of the red liquid, their judgments were not simply based on how the red liquid and blooming covaried. Instead, participants systematically used observations in which the alternative cause (blue liquid) was held constant (a strategy referred to as *conditionalizing*), just as scientists control for potential confounding variables in experimental design (see also Goodie, Williams, & Crooks, 2003; Waldmann & Hagmayer, 2001). Conditionalizing is advantageous because it often prevents wrongly attributing causal efficacy to a noncause. For instance, upon observing that more men than women are scientists, one should eliminate differences in socialization before concluding genetic differences as the cause (see also Simpson's paradox; Simpson, 1951).

Although conditionalizing may be useful, the strategy is often not feasible because it requires alternative causes to be observed. Sometimes alternative causes are unobserved because they require special instruments or methods to be observed (e.g., genetic influences on cancer). More frequently, learners do not collect observations about alternative causes simply because they are unable to consider all possible alternative causes of a particular event. Thus, it seems that learners must constantly deal with unobserved alternative causes. Though all modern theories of causal learning incorporate some mechanism for conditionalizing, very few attempt to actively deal with unobserved alternative causes, as we illustrate next.

Mechanisms for Unobserved Cause Learning

In this section, we review how existing models of causal learning deal with unobserved causes. For purposes of comparison, we

Christian C. Luhmann and Woo-kyoung Ahn, Department of Psychology, Yale University.

Portions of this article form the basis of Christian C. Luhmann's doctoral dissertation and have been presented previously at the Annual Conference of the Cognitive Science Society (August 2003, Boston, Massachusetts; July 2006, Vancouver, British Columbia, Canada). This project was supported by National Institute of Mental Health Grant RO1 MH57737 to Woo-kyoung Ahn.

We thank Gordon Logan, Thomas Palmeri, and David Noelle for useful feedback on drafts of this article.

Correspondence concerning this article should be addressed to Christian C. Luhmann, P.O. Box 208001, New Haven, CT 06520. E-mail: christian.luhmann@yale.edu

		Effect: Present	Effect: Absent
Cause	Present	A	B
	Absent	C	D

Figure 1. A contingency table summarizing the covariation between two binary events. Each cell of the table represents one of the possible observations.

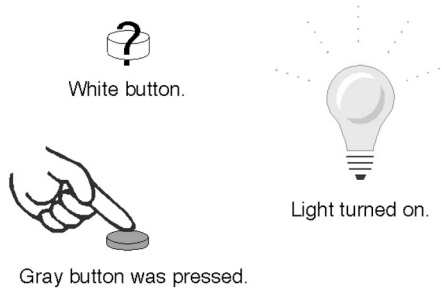

Figure 2. A sample trial. The states of the gray button and the light are observed on every trial. The white button is unobserved; information about its state is unavailable on every trial.

have segregated these existing models along two critical dimensions (see Table 1). The first dimension, represented by the rows of Table 1, captures the extent to which a model makes inferences about unobserved alternative causes. Models can range from remaining entirely agnostic about strength of an unobserved alternative cause to making full inferences about it. The second dimension, represented by the columns of Table 1, represents whether computations of causal strengths take place on each trial of observation (iterative) or only at the end of all observations (simultaneous).

To illustrate how models varying along these two dimensions deal with unobserved alternative causes, we will consider a relatively simple situation that includes one observed cause, one unobserved cause, and a single effect, each taking on one of two values (present or absent). See Figure 2 for an example trial and Figure 3 for a sample covariation table for this situation.

ΔP

ΔP is a method of computing contingency. It does not include any mechanism for dealing with unobserved causes and performs its computation all at once at the end of all observations (see Table 1). ΔP is the probability of the effect occurring in the presence of the observed cause minus the probability of the effect occurring in the absence of the observed cause (e.g., $[10 / (10 + 10)] - [10 / (10 + 10)] = 0$ for Figure 3).

Applying the same computation to an unobserved cause is simply not possible because the probabilities required to compute ΔP are unavailable for unobserved causes. One might notice that if we assume that there is only one unobserved cause, then it must be present when the effect occurs in the absence of the observed cause (i.e., during the 10 occasions represented in the lower left quadrant of Figure 3). Unfortunately, even with this insight, the causal strength of the unobserved cause cannot be computed because the presence/absence of the unobserved cause is unconstrained during the remaining 30 observations. At one extreme, the presence of the unobserved cause could correlate perfectly with the effect on the remaining 30 occasions (i.e., the unobserved cause would be present in all 10 cases in which the effect and the observed cause were present, and absent in all 20 cases in which the effect was absent), resulting in $\Delta P = 1.0 - 0.0 = 1.0$. At the other extreme, the presence of the unobserved cause could negatively correlate with the effect on the remaining 30 occasions (i.e., the unobserved cause would be present on all 20 occasions when the effect was absent, and absent on 10 of the occasions when the effect and the observed cause were present), resulting in $\Delta P = .5 - 1.0 = -0.5$. Thus, ΔP cannot provide a unique estimate for the causal strength of the unobserved cause.

Table 1
Summary of Potential Unobserved Causal Learning Mechanisms

	Operation	
Unobserved cause behavior	Simultaneous	Iterative
Unknown unobserved cause behavior	ΔP (Jenkins & Ward, 1965) Power PC (Cheng, 1997) Probability of sufficiency (Pearl, 2000)	Rescorla & Wagner (1972) Danks et al. (2003) (assuming more than one unobserved cause)
Static unobserved causes	Power PC (Cheng, 1997) Probability of sufficiency (Pearl, 2000) (amended with $P(U = 0.5)$)	Rescorla–Wagner Danks et al. (2003) (assuming a single unobserved cause)
Dynamic unobserved causes	EM (Dempster et al., 1977)	BUCKLE

Note. The rows describe how different models deal with unobserved causes. Those in the top row either ignore unobserved causes or are otherwise unable to estimate the strength of any given unobserved cause. Models in the middle row are models that can produce unobserved cause estimates when assuming that the unobserved cause occurs with some constant probability. Models in the bottom row allow unobserved causes to vary from trial to trial. The two columns divide the models into those that perform a single computation at the end of the trial sequence (simultaneous) and those that perform computations on every trial (iterative). EM = expectation maximization algorithm; BUCKLE = bidirectional unobserved cause learning.

Observed Cause	Effect: Present	Effect: Absent
Present	10	10
Absent	10	10

Figure 3. A summary of 40 observations characterizing the situation presented in Figure 2. The contingency table summarizes the covariation of the observed cause and the effect. No such table can be fully constructed for an unobserved cause.

Power Models

The next approach we consider is embodied by a class of models that includes the causal power theory of the probabilistic contrast model, or the power PC theory (Cheng, 1997), and Pearl's probability of sufficiency (Pearl, 2000). With sophisticated use of probability theory and critical assumptions, these models attempt to provide causal strength estimates, as opposed to mere covariation, such as ΔP. Thus, we refer to this class of models as power models. The power models acknowledge the existence of unobserved alternative causes but cannot infer their properties. The power models, like ΔP, perform their computations in a single step (see Table 1).

Take the power PC theory (Cheng, 1997) as an example. This theory suggests that q_O, the causal strength of the observed cause, O, in a situation depicted in Figure 2 is equivalent to Equation 1.

$$q_O = \frac{\Delta P}{1 - P(U = 1) \cdot q_U} \quad (1)$$

where $P(U = 1)$ is the base rate of an unobserved cause, U, and q_U is the causal power of the unobserved cause.

The immediate problem with this equation is that the denominator contains two unknowns: Both the strength and the base rate of the unobserved cause are unavailable. To deal with this problem, the power PC theory assumes that Equation 1 is used in situations in which unobserved alternative causes occur independently of the observed cause. Under this assumption, $P(U = 1) \cdot q_U$ becomes equivalent to the probability of the effect given the presence of the observed cause, which is an observable quantity, and the causal strength of the observed cause can be computed (see Cheng, 1997, for the proof). Thus, this assumption eliminates any need to make inferences about strengths of unobserved alternative causes.

There are two remaining problems with this treatment in relation to the current study. First, recent studies have demonstrated that even when they are willing to make causal judgments, people do not necessarily assume that unobserved alternative causes occur independently of the observed cause (see Hagmayer & Waldmann, 2007; Luhmann, 2005). Experiment 3 in this article revisits this issue. Second, even when alternative causes are independent, there is no unique solution for the strength of the unobserved cause (q_U) because $P(U = 1)$ is still unknown. This ambiguity prevents the power PC theory (and probability of sufficiency; see Pearl, 2000) from making firm predictions about the strength of unobserved causes. In contrast, when Luhmann and Ahn (2003) presented participants with a series of trials similar to the one shown in Figure 2, all participants were willing to provide causal strength judgments of an unobserved cause, even when explicitly given the option not to do so. Because $P(U = 1)$ was unknown to participants in this experiment, the power models would not be able to produce unique estimates and thus cannot account for people's willingness to judge the unobserved cause (q_U).

As we will show when reviewing other models, the problem plaguing the power models is pervasive. The strength of unobserved causes is underdetermined without additional information about how or when the unobserved cause is present or absent. Acknowledging this problem suggests one relatively simple solution for the power models to compute the strength of unobserved causes and to potentially explain people's willingness to provide judgments. If learners made simplifying assumptions about $P(U = 1)$ (e.g., $P(U = 1) = .5$, meaning unobserved causes are present on about half of the trials), then there would be only one unknown to be solved for: q_U. Thus, by simply assuming that the unobserved causes occur with some fixed probability, the power models are now able to provide the strength estimates people are able to provide.[1] These amended power models occupy a new cell in Table 1 to acknowledge the psychological assumptions about the occurrence of unobserved causes that must be added (and tested).

To summarize, the power models cannot provide q_U and thus fail to account for the full range of learning behavior (e.g., Luhmann & Ahn, 2003). However, with a simple amendment of fixing $P(U = 1)$, they can estimate q_U and thus have the potential to explain people's judgments. These models are thoroughly tested below.

Rescorla–Wagner Model

The learning model described by Rescorla and Wagner (1972; hereafter R-W) is a classic iterative learning model (and thus occupies the right column in Table 1) that actually mirrors the predictions of ΔP under many conditions (Cheng, 1997; Danks, 2003; Shanks, 1995). The model assigns each cause and each effect a node in a simple network. According to R-W, learning amounts to adjustments of strengths according to Equation 2.

$$\Delta q = \alpha\beta(\lambda - \Sigma q) \quad (2)$$

In this equation, λ is an indicator of whether the effect is present ($\lambda = 1$ in our simulations) or absent ($\lambda = 0$). The parameters α and β are the saliency of the cause and the effect, respectively. We will assume a value of .5 for β and will fit the value of α using procedures described below. The parenthetical quantity is the amount of error. This error is computed as the difference between the observed effect (λ) and the predicted value of the effect (Σq; the summed strengths of the causes present on that occasion). The resulting quantity, Δq, is then used to adjust the strength of each cause present on that occasion. Causes that are absent never have their strengths adjusted.

On its surface, R-W does not appear to deal with unobserved causes at all. However, there is one algorithmic detail that may allow R-W to produce unobserved cause judgments. R-W always

[1] We thank Tom Griffiths for suggesting this strategy.

includes a single additional input node to represent the experimental context or background. This background is essentially the set of all unobserved causes. For example, Shanks (1989) stated that "occurrences of the [effect] in the absence of the target cause . . . must be attributed to the background" (p. 27). Because the experimental context is present on all trials, R-W assumes that this cue is also constantly present and its strength is updated just as any other cause. Now, the unobserved cause depicted in Figure 2, the unobserved cause that participants are asked to judge, is certainly part of the experimental context. However, there is no way to extract the strength of any single unobserved cause from the learned strength of the experimental context (a potentially infinite set of such causes). This is why it is placed in the top row of Table 1. Thus, this aspect of R-W, though suggestive, does not actually help it in explaining whether and how people make inferences about unobserved causes (Luhmann & Ahn, 2003).

However, as with the power models, amendments can be made that allow R-W to estimate the strength of the unobserved cause. Namely, the strength of the experimental context can be used to directly judge the single, unobserved, alternative cause. Conceptually, this assumption is equivalent to ignoring all aspects of the context except the unobserved cause of interest, which, for human participants, may not be that much of a stretch.[2] This assumption also predicts that the unobserved cause is treated as though it were present on every trial (and thus occupies the middle row of Table 1). This brings the R-W model in line with the amended power models; both operate with an unobserved cause occurring with a fixed probability (.5 for the amended power models or 1 for R-W).

Power Variant of R-W

Whereas the original R-W model can be thought of as a trial-by-trial, algorithmic variant of ΔP, Danks, Griffiths, and Tenenbaum (2003) have described a learning process that computes causal power (Cheng, 1997). Their proposal is highly similar to the traditional R-W model in that it uses a trial-by-trial error-correction algorithm and assumes that an unobserved cause is constantly present and that its strength is adjusted just like observed causes. There are two modifications that allow the power variant of R-W to compute the causal power.

First, whereas R-W assumes that causal strengths combine additively to produce their effects, the power variant of R-W assumes that generative causes combine in the manner of a noisy-OR gate and preventative causes combine in the manner of a noisy-AND-NOT gate (e.g., Cheng, 1997; Griffiths & Tenenbaum, 2005; Novick & Cheng, 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). For example, if two generative causes (A and B) were present on some occasion, R-W would simply add the two strengths (i.e., $P(E) = q_A + q_B$) to predict the probability of the effect occurring. Under the noisy-OR assumption, the effect occurs to the extent that one or the other cause is sufficient to produce it (illustrated in Equation 3).

$$P(E) = q_A + q_B - (q_A \cdot q_B) \quad (3)$$

Second, whereas R-W never adjusts the strength of causes that are absent, the power variant of R-W adjusts the strength of all causes on each occasion. Present causes are adjusted just as in R-W. Strengths of absent causes are increased when the effect is predicted to be present but is actually absent, and decreased when the effect is predicted to be absent but is actually present. To accomplish this, the value of α in Equation 2 is positive when the cause is present (as it was in traditional R-W) and negative when the cause is absent.

When it comes to explaining unobserved cause learning, the power variant of R-W is in the same situation as traditional R-W. If the constantly present alternative cause is treated as the set of all unobserved, alternative causes, it is unable to produce strength estimates of any single unobserved cause. However, if the constantly present alternative cause is treated as representing the single unobserved cause of interest, then it may be able to explain participants' judgments.

Expectation Maximization

The last models we consider make more sophisticated inferences about unobserved causes. These models actually estimate (rather than assume) the value of $P(U = 1)$ using the available data. Because the values of the observed variables change from trial to trial, inferences about the unobserved cause also vary from trial to trial (see the bottom row of Table 1).

A standard method to accomplish such learning in the field of statistics is to apply the expectation maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). The EM algorithm alternates between computing the most likely parameter values and filling in the missing data on the basis of the new parameter estimates. These two steps are repeated until the estimates converge. Thus, EM would attempt to simultaneously estimate the causal strengths and the missing data (i.e., the value of U on every trial).

The EM algorithm was not proposed as a psychological model of causal learning (though see Fried & Holyoak, 1984). We will not provide detailed discussion, but it is worth noting that this algorithm is unable to arrive at unique estimates of $P(U = 1)$ and q_U. EM has been shown (Dempster et al., 1977) to converge on the maximum likelihood estimate (MLE), which, in this case, is equivalent to the power models (see Griffiths & Tenenbaum, 2005). Given this equivalence, it is no surprise that there exist multiple sets of parameter values that are equally consistent with the data summarized in Figure 3. For example, the unobserved cause could have a strength of 1 (i.e., completely sufficient to produce the effect) but be present on only half of the trials—that is, $P(U = 1) = .5$—or it could have a strength of .5 and be present on all trials—that is, $P(U = 1) = 1$. Though it is possible to use EM (or MLE) to compute estimates for more specific quantities—for example, $P(U = 1|E)$ or the value of $P(U = 1)$ for a particular trial—doing so would only produce larger sets of equally likely parameter estimates.

By surveying the entire data set at once, the EM algorithm (and MLE) is forced to contemplate all possible combinations of parameter values. This method of operation allows it to definitely find the most likely set of parameter values, but it also causes increased ambiguity, as illustrated above. Our own model, BUCKLE, uses the principles behind EM but learns in an iterative manner. As we will show, BUCKLE is able to produce estimates

[2] Indeed, in the current Experiment 1, we explicitly instructed participants that there was only a single alternative cause for the effect, so this assumption was satisfied in that experiment.

of both the strength and presence/absence of unobserved causes. We discuss the similarities between EM and BUCKLE in the General Discussion.

BUCKLE

As we illustrated above, no set of estimates—that is, q_O, q_U, and $P(U = 1)$—uniquely describes the covariation between an observed cause and its effect. To overcome this problem, BUCKLE uses a strategy of alternating between estimating the missing data—for example, $P(U = 1)$—and estimating the parameters of interest (i.e., q_O and q_U, much like the EM algorithm described above) with the added psychological constraint that data be processed in an iterative manner. Specifically, we argue that learners infer how likely the unobserved cause is to be present as each observation is encountered and then proceed with learning as if the presence/absence of the unobserved cause were available in the input. Here, we first generally describe the model before presenting a formal description.

BUCKLE is a trial-by-trial learning model and operates using two steps, each of which is performed as each observation is encountered. Figure 4 provides a schematic of BUCKLE's operation. BUCKLE first computes the probability that the unobserved cause is present on the current trial, removing the uncertainty about the occurrence of the unobserved cause. BUCKLE computes the probability that the unobserved cause is present on the basis of the strength of each cause and the presence/absence of the observed causes and effect. The strengths used are those computed on the previous trial. To illustrate this step, suppose a learner, on the basis of previous trials, believes that an observed cause is very weak. If this learner now observes a trial in which both the observed cause and the effect are present, she would infer the unobserved cause to be more likely present if she believes the unobserved cause is strong rather than weak (see the next section for formal algorithms).

Once this probability of the unobserved cause is determined, BUCKLE then revises its estimates of the causal strengths of both the observed and unobserved causes in the second step. This step requires, among other things, information about the probability of the unobserved cause being present on this occasion, which is provided by the first step of that trial. The strength adjustment is accomplished via an error-correction algorithm like that used by the R-W variants described above. BUCKLE then uses the updated causal strengths on the next trial when inferring how likely the unobserved cause is to be present and updating each cause's strength. On each trial, these two steps repeat.

Thus, BUCKLE does not assume that the unobserved cause is constantly present. However, it also avoids the difficulties associated with determining both the occurrence and the strength of unobserved causes. The critical difference is that, unlike the simultaneous models (the left column in Table 1), BUCKLE does not attempt to compute the value of these two quantities simultaneously. Instead, BUCKLE alternates between (a) dealing with the occurrence of the unobserved cause and (b) adjusting the strength of the unobserved cause. During each computation to estimate one of these quantities, the value of the other quantity is known and static.

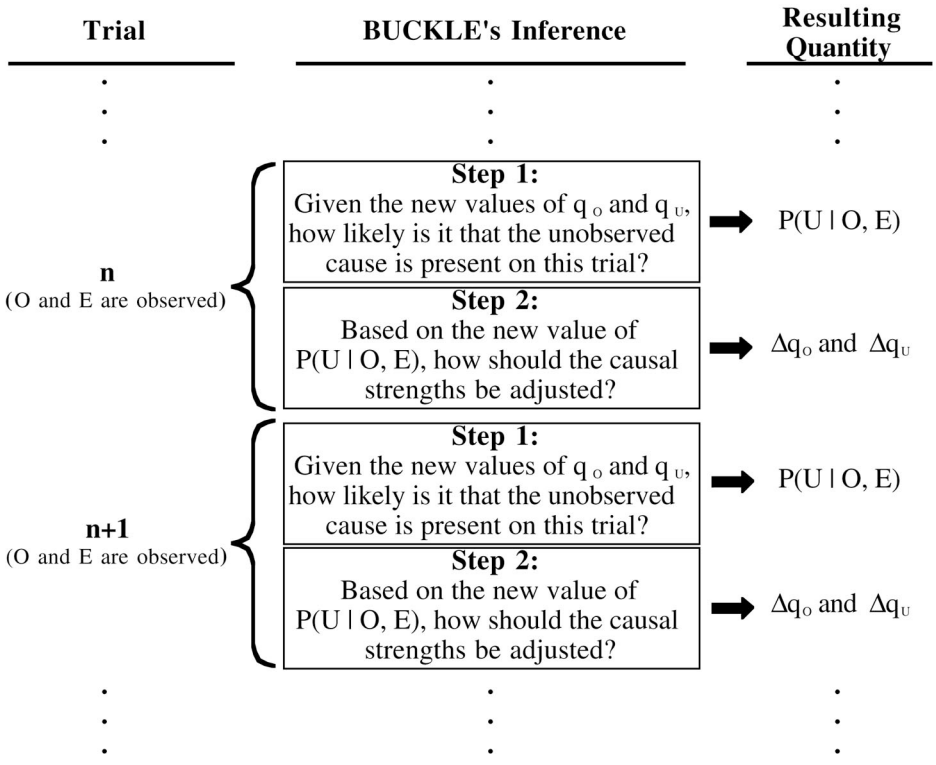

Figure 4. A diagram illustrating the operation of the two steps of BUCKLE (bidirectional unobserved cause learning). *O* represents a single observed cause, *U* represents a single unobserved cause, and *E* represents a single effect.

To summarize, the most critical aspect of BUCKLE is that it infers how likely the unobserved cause is to be present as each observation is encountered and that this likelihood estimate is used in updating causal strengths of unobserved and observed causes. BUCKLE's process is highly similar to the generalized EM algorithm described by Dempster et al. (1977), which is guaranteed to converge on a locally optimal solution, except that BUCKLE adds the psychological constraint that computations must be performed on a single observation at a time in an iterative fashion (see the General Discussion for more on the similarities between these algorithms).

Formal Description of BUCKLE

The formal details of BUCKLE provided here are for the simple case illustrated in Figure 2, which includes a single observed cause (O), a single unobserved cause (U), and a single effect (E).[3] Using the conventional notation, 1 represents presence and 0 represents absence (e.g., $O = 1$, $E = 0$ for the present observed cause and absent effect). For the sake of brevity, we will often use the conventional abbreviation in the text (e.g., $O\bar{E}$ for an occasion on which the observed cause is present and the effect is absent). The strengths of O and U, specifically their causal sufficiency (just as with the power models), will be represented by q_O and q_U, respectively.

Step 1: Inference of unobserved cause. The first step taken by BUCKLE is to infer how likely it is that the unobserved cause is present in a given trial. To do this, the values of O and E are first set according to the current states of the observed cause and effect in the current observation (i.e., $O = o$, $E = e$). The probability that the unobserved cause is present is then computed using Bayes's theorem.

$$P(U = 1|O = o,E = e) = \frac{P(E = e|O = o,U = 1) \cdot P(U = 1|O = o)}{\sum_{u=[0,1]} P(E = e|O = o,U = u) \cdot P(U = u|O = o)} \quad (4)$$

We will assume that the prior probability of the unobserved cause occurring is always .5—that is, $P(U = 1|O = 1) = P(U = 1|O = 0) = .5$ (see the General Discussion for more on this assumption). This allows us to simplify Equation 4.[4]

$$P(U = 1|O = o,E = e) = \frac{P(E = e|O = o,U = 1)}{\sum_{u=[0,1]} P(E = e|O = o,U = u)} \quad (5)$$

To compute $P(E = e \mid O = o, U = u)$, we need to specify how causes combine to influence their effects. As in the power models, we will assume that causes combine in the manner of a noisy-OR gate when generative and a noisy-AND-NOT gate when preventative (this can be changed depending on the specific situation; see the General Discussion). Thus, when O and U are generative (i.e., $q_O, q_U > 0$), the effect occurs according to Equation 6.

$$P(E = 1|O = o,U = u) = (O \cdot q_O) + (U \cdot q_U) - [(O \cdot q_O) \cdot (U \cdot q_U)] \quad (6)$$

For example, if both O and U are present, Equation 6 becomes $P(E = 1|O = 1,U = 1) = q_O + q_U - q_O \cdot q_U$. If O is present and U is absent, Equation 6 becomes $P(E = 1|O = 1,U = 0) = q_O$.

When U is preventative and O is generative (i.e., $q_U < 0$, $q_O > 0$), E occurs according to Equation 7.

$$P(E = 1|O = o, U = u) = O \cdot q_O \cdot \{[U \cdot (1 + q_U)] + (1 - U)\} \quad (7)$$

When U is generative and O is preventative (i.e., $q_U > 0$, $q_O < 0$), E occurs according to Equation 8.

$$P(E = 1|O = o, U = u) = U \cdot q_U \cdot \{[O \cdot (1 + q_O)] + [1 - O]\} \quad (8)$$

When neither cause is generative, $P(E = 1 \mid O = o, U = u) = 0$.

Combining these individual expressions allows BUCKLE to compute the probability of U. For example, imagine a trial on which q_O and q_U are believed to be positive (i.e., O and U produce rather than prevent their effects) and on which both the observed cause and the effect occur (i.e., $O = 1$, $E = 1$). Because q_O and q_U are both positive, BUCKLE uses Equation 6 to expand Equation 5, which results in Equation 9.

$$P(U = 1|E = 1,O = 1) = \frac{q_O + q_U - q_O \cdot q_U}{q_O + (q_O + q_U - q_O \cdot q_U)} \quad (9)$$

As another example, imagine that the effect is still present (i.e., $E = 1$) and O is still generative (i.e., $q_O > 0$) and present on a given trial (i.e., $O = 1$) but that U is now preventative ($q_U < 0$). BUCKLE would use Equation 7 to expand Equation 5, which would result in Equation 10.

$$P(U = 1|E = 1,O = 1) = \frac{q_O \cdot (1 + q_U)}{q_O + [q_O \cdot (1 + q_U)]} \quad (10)$$

Thus, despite the fact that the observation (i.e., O and E) is the same as above, Equation 5 will be computed differently because the current belief about U's influence is different from the previous example. The Appendix provides the computations for all eight possible cases. BUCKLE decides which of these eight expressions to use dynamically, according to the current values of O and E (available from the input) and q_U and q_O (obtained from the previous trial). Because BUCKLE determines the influence (generative vs. preventative) of each cause according to the input, it is possible to begin a trial sequence with a positive value of q_U and end that same trial sequence with a negative value of q_U. BUCKLE would compute the probability of the unobserved cause using one expression (e.g., Equation 9) on trials where q_U was positive and a different expression (e.g., Equation 10) on trials where q_U was negative.

[3] Typical real-world situations include multiple observed causes. People may acknowledge this fact, in which case BUCKLE could be modified to accommodate additional unobserved causes. On the other hand, people may lump all of the unobserved alternative causes into a single composite cause (e.g., Cheng, 1997), in which case it would be more appropriate to use the version described here. Regardless, different causal situations are simply generalizations of what is described here.

[4] See the Appendix for the complete derivation, including details about how undefined quantities are handled.

Step 2: Learning algorithm. The second step of BUCKLE is to adjust the strength of each causal relationship (i.e., q_U and q_O), using, among others, the inferred value of $P(U = 1)$ obtained from Step 1. To do so, BUCKLE uses an error-correction algorithm much like R-W (with a few minor changes, noted below). BUCKLE first makes a prediction about the occurrence of E based on the presence/absence of O and U and the current estimates of their strengths. The discrepancy between this prediction and the actual, observed occurrence of the effect is used to adjust the strength estimates.

BUCKLE makes its prediction about the occurrence of the effect as described above (e.g., Equation 3). We take BUCKLE's prediction about the occurrence of the effect to be equal to Equation 11.

$$E_{\text{predicted}} = \{P(E = 1|O = o,U = 1) \cdot P(U = 1|O = o,E = e)\} + \{P(E = 1|O = o,U = 0) \cdot [1 - P(U = 1|O = o,E = e)]\} \quad (11)$$

BUCKLE's prediction,[5] $E_{\text{predicted}}$, is then compared with the actual value of E, and the difference (i.e., error) is used to adjust the strength estimates, as in Equations 12 and 13.

$$\Delta q_O = \alpha_O \beta (E - E_{\text{predicted}}) \quad (12)$$

$$\Delta q_U = \alpha_U \beta (E - E_{\text{predicted}}) \quad (13)$$

The strength of each cause is updated separately. The quantities α and β represent learning rates associated with causes and effects, respectively. A value of .5 will be used for β (see Appendix Table A1). When the observed cause is present, $\alpha_O = \alpha_{O\text{-present}}$, where $\alpha_{O\text{-present}}$ will be treated as a free parameter and allowed to vary between 0 and 1. When the observed cause is absent, $\alpha_O = 0$. For the unobserved cause, the value of α is computed using Equation 14, which takes into account the fact that the unobserved cause is present only with some probability.

$$\alpha_U = \alpha_{U\text{-present}} \cdot P(U = 1|O = o,E = e) \quad (14)$$

The variable $\alpha_{U\text{-present}}$ will be treated as a second free parameter and allowed to vary between 0 and 1. Equation 14 results in $\alpha_U = 0$ when $P(U = 1 \mid O = o, E = e) = 0$, and $\alpha_U = \alpha_{\text{U-present}}$ when $P(U = 1 \mid O = o, E = e) = 1$, just as for the observed cause. For values of $P(U = 1 \mid O = o, E = e)$ between 0 and 1, α increases linearly and in proportion to the value of $P(U = 1 \mid O = o, E = e)$.

To review, BUCKLE completes two steps for each observation. BUCKLE first infers how likely the unobserved cause is to be present and then adjusts the causal strengths of all present causes. Beyond these two steps, the particular algorithms behind each step of BUCKLE's operation are interchangeable (see the section *BUCKLE's Assumptions* in the General Discussion for more on this point).

Simulation of BUCKLE on Observed Cause Learning

Although BUCKLE's innovation lies in unobserved cause learning, the first order of business is to ensure that BUCKLE can replicate people's judgments of observed causes. We selected Experiment 3 of Buehner, Cheng, and Clifford (2003) for a test case. Their participants received 10 different contingency conditions, each consisting of 24 randomly ordered trials depicting the presence or absence of the cause and effect. At the end of each condition, participants judged the likelihood that an effect would occur given that the observed cause was present. This experiment is most appropriate for testing BUCKLE because (a) its dependent measure elicits the quantity computed by BUCKLE; (b) it uses trial-by-trial learning procedures, which is the way BUCKLE updates its beliefs; and (c) it uses 10 different contingency conditions, allowing for generality of the tests.

We ran BUCKLE in each of the 10 conditions from this experiment. Because BUCKLE is sensitive to trial order (see Experiment 4), we created 1,000 simulated participants, with a randomized order of the observations in each condition. Using the directed search algorithm described by Hooke and Jeeves (1960), BUCKLE's $\alpha_{O\text{-present}}$ and $\alpha_{U\text{-present}}$ parameters were fitted for each simulated participant to the mean causal judgments reported by Buehner et al (2003). All other parameters were set as described in the Appendix (Table A1).

We calculated the fit between Buehner et al.'s (2003) means from the 10 conditions in Experiment 3 and the mean strength estimates generated by BUCKLE from the 1,000 simulated participants for each of the same 10 contingency conditions. BUCKLE accounted for 98% of the variance in actual participants' judgments. This fit is as good as that of the power PC theory itself ($R^2 = .97$) and better than that of ΔP ($R^2 = .87$).[6] This result

[5] Because the value of $P(U = 1 \mid O = o, E = e)$ is influenced by the value of E, it may not be obvious that Equation 11 can produce values of $E_{\text{predicted}}$ that do not match the actual value of E. However, imagine that a learner believes that $q_O = 0$ and $q_U = 0$ and then observes a trial on which $O = 1$ and $E = 1$. The first step will produce $P(U = 1 \mid O = 1, E = 1) = .5$. The value of $E_{\text{predicted}}$ would then be 0 because $q_O = 0$ and $q_U = 0$, and this prediction would be incorrect (because E was present). The bottom line is that just because the value of $P(U = 1 \mid O = o, E = e)$ can be shifted according to the observed data, it cannot always (and generally will not) take on a value that leads to perfect predictions.

[6] People's causal strength judgments in the Buehner et al. (2003) study, as well as BUCKLE's simulation of this experiment, were probabilistic (e.g., .5) rather than deterministic (i.e., 0, 1, or –1). One might wonder whether such probabilistic estimates contradict Luhmann and Ahn's (2005) claim that causal power is deterministic. Note that the earlier claim about deterministic causes pertains exclusively to Cheng's notion of causal power rather than causal sufficiency, which BUCKLE computes (which is akin to Cheng's [2000] notion of contextual causal power or Pearl's [2000] probability of sufficiency). Computation of causal power requires several conditions to be met. For example, there may not be any alternative, preventative causes present during learning (see Cheng, 1997, for the full set). As discussed in Luhmann and Ahn (2005), if and only if participants accept this entire set of requirements will estimates be deterministic. As Luhmann and Ahn further discussed, however, these assumptions are extremely unlikely to be satisfied in the real-world situations as well as the scenarios used in Buehner et al.'s (2003) experiment. For instance, participants could have believed that the ostensible causes (medications) did not impinge directly on their effects but rather operated through a series of intermediate causal links (perhaps via the shorthand belief in some unnamed mechanism). We (Luhmann & Ahn, 2005) have argued that participants with such beliefs must have computed a more general measure of causal sufficiency (as BUCKLE does) rather than the power PC theory's specific version of causal power. When these requisite assumptions for causal power are dropped, the resulting causal sufficiency may be probabilistic. For instance, the probabilistic estimates of causal sufficiency would reflect indeterminacy that results from an intervening mechanism.

should not be too surprising given the previous success of the related model by Danks et al. (2003).

Experiments on Unobserved Cause Learning

The above simulation showed that BUCKLE is able to account for a significant portion of people's behavior in learning of observed causes. However, the crux of BUCKLE is to explain learning of unobserved causes and its impact on observed causal learning. At this point, there are hardly any data on learning of unobserved causes. Below, we report five sets of experiments on unobserved causal learning and test BUCKLE against other causal learning models. The models in the first row of Table 1 assume that people cannot estimate strengths of unobserved causes, and so Experiments 1 and 2 examine whether learners do in fact learn about unobserved causes. Experiment 3 investigates whether people treat unobserved causes as constantly present, as suggested by the amended power and R-W models (i.e., models in the middle row of Table 1), or whether people infer how likely the unobserved cause is to be present on each trial, as suggested by BUCKLE. Experiment 4 investigates the iterative nature of BUCKLE by examining whether people are sensitive to the sequence in which trials are presented. Experiment 5 further explores an apparent inconsistency between the current findings and previous findings from the developmental literature.[7]

Experiment 1: Judging Unobserved Causes

In this experiment, participants were presented with a situation that included a single observed cause, a single unobserved cause, and a single effect (e.g., Figure 2). Four representative contingencies between the observed cause (O) and the effect (E) were used, as illustrated in Figure 5. In the *Perfect* and the *Zero* conditions, the correlations between O and E were 1 and 0, respectively. The remaining two conditions represent moderate, probabilistic relationships between O and E. In the *Unnecessary* condition, O was unnecessary but sufficient for E, and in the *Insufficient* condition, O was insufficient but necessary for E. In each condition, participants observed the contingency between O and E specified in Figure 5 in a trial-by-trial manner and, after observing all trials in a condition, provided causal strength estimates for both the observed and the unobserved causes.

One of the main goals of Experiment 1 was to test the prediction derived from ΔP and the power models that learners would be unable to provide any unique solution for strengths for the unobserved causes. As mentioned earlier, we have previously shown (Luhmann & Ahn, 2003) that participants are willing to provide judgments of an unobserved cause even when explicitly given the option not to do so. However, willingness alone is weak evidence for unobserved cause learning. Learners still might not be able to provide systematic judgments of unobserved causes (e.g., no differences between contingencies), in which case we would conclude that learners, like the power models, are unable to make unique judgments about the strength of unobserved causes.

However, if participants do provide systematic judgments of the unobserved cause, their estimates can be compared with predictions of BUCKLE, R-W, and the power variant of R-W. Because these models' parameters should be fit to participants' data, the models' quantitative predictions are described after the report of the results. Here, we briefly address only BUCKLE's predictions conceptually (see Figure 5).

In the Unnecessary condition, BUCKLE increases the strength of the unobserved cause because there are several observations on which the unobserved cause certainly co-occurs (with the effect $\bar{O}E$). Because BUCKLE assumes that causes compete with each other to accrue strength, the increased strength of the unobserved cause limits the strength of the observed cause from reaching ceiling levels (cf. power PC). In the zero condition, the unobserved cause again gains strength because of the $\bar{O}E$ observations. The observed cause does not generally accrue any strength because of the zero correlation (as predicted by both ΔP and the power PC theory). In the Perfect condition, BUCKLE predicts that the observed cause will accrue nearly all of the causal strength because of the perfect correlation and the unobserved cause will accrue almost none (as predicted by both ΔP and the power PC theory). In the Insufficient condition, the strength of the observed cause reaches moderate levels because its covariation suggests moderate sufficiency (as predicted by both ΔP and the power PC theory). The strength of the unobserved cause also reaches moderate levels because the observed cause that has only moderate causal strength is unable to completely explain the occurrence of the effect. In short, BUCKLE predicts the unobserved causal strength to be higher in the Unnecessary and the Zero conditions than in the Insufficient and the Perfect conditions.

The design of Experiment 1 also allows us to examine the well-known discounting principle (Kelley, 1972). Although not presented as a full computational model in this article (but see Luhmann, 2006, for tests of a constraint satisfaction model that instantiate this principle), the discounting phenomenon has been shown to be robust, and thus it is a highly plausible heuristic for the current situation. According to this principle, when a strong cause is observed, the alternative unobserved cause will be discounted (i.e., inferred to be weak), and vice versa. Taking ΔP as our measure of causal strength, the observed cause is equally strong in the Unnecessary and Insufficient conditions, and therefore, the discounting principle predicts no difference between these two conditions in terms of the unobserved causal strength. However, BUCKLE predicts the unobserved causal strength to be lower in the Insufficient than in the Unnecessary condition (see Figure 5). If we take the power models as our measure of causal strength, assuming all of the requisite assumptions are met, then the observed cause is equally strong in the Unnecessary and Perfect conditions ($q_O = 1$), and again the discounting strategy predicts no difference between these two conditions in terms of the unobserved causal learning. However, BUCKLE predicts the unobserved causal strength to be lower in the perfect condition than in the unnecessary condition.

We conducted two experiments, Experiments 1A and 1B, differing only in the dependent variables, because different models measure different quantities. Experiment 1A measured causal sufficiency (i.e., the degree to which a cause is sufficient to bring about its effect; see Buehner et al., 2003) to test BUCKLE and the

[7] In all experiments, participants were undergraduate students at either Vanderbilt University or Yale University, participating for partial fulfillment of course credit or for pay. Details of the experiments reported in this article can be found in Luhmann (2006).

	Condition			
	Unncessary	Zero	Perfect	Insufficient
Contigency Structure	O: 7, 0 / $\overline{O}$: **7**, 7	O: 7, **7** / $\overline{O}$: **7**, 7	O: 7, 0 / $\overline{O}$: 0, 7	O: 7, **7** / $\overline{O}$: 0, 7
BUCKLE's Predictions				
Observed Cause	Moderate	Low	High	Moderate
Unobserved Cause	High	High	Low	Moderate
ΔP				
Observed Cause	50	0	100	50
Power PC				
Observed Cause	100	0	100	50
Unobserved Cause (assuming P(U)=.5)	100	100	0	0

Figure 5. The design used in Experiment 1. Each condition contains OE and $\overline{OE}$ observations. Only the presentation of $\bar{O}E$ and $O\bar{E}$ observations differs, as shown in bold. BUCKLE = bidirectional unobserved cause learning.

power models. Experiment 1B used the traditional but ambiguous method of eliciting causal judgments (e.g., To what extent does X cause Y?) typically used when evaluating ΔP and R-W (e.g., Shanks, 1987).[8]

Method

Stimuli consisted of four electrical systems, each containing one button whose state (pressed or not) was observable, one button whose state was unobservable and a single light. Information about the behavior of these systems was presented visually on a computer (see Figure 2, for an example). The unavailable state of the unobserved button was denoted with a large question mark superimposed over the button. The state of the light (on or off) was always observable. Figure 5 shows the cell frequencies for each condition.

Participants ($N = 24$ in Experiment 1A; $N = 30$ in Experiment 1B) first received general instructions about what they would observe (e.g., two buttons, a light) and what the symbols meant (e.g., question mark). They were also told that nothing other than the two buttons could influence the light. This was done to equate participants' assumptions about the situation with the assumptions used in the modeling reported later. (See Experiment 2 for alternative instructions.) Each participant saw all four systems in a counterbalanced order. The trials within each system were presented in a quasi-randomized order to evenly distribute the different types of trials.

After viewing the entire set of trials, participants rated the causal strength of the observed and unobserved buttons separately. In Experiment 1A participants were told, "Imagine running 100 new tests in which the [color] button was pressed and the [color] button was not. On how many of these tests do you expect the light to turn on?" Participants responded with a number between 0 and 100. In Experiment 1B, participants were asked to "judge the extent to which pressing the [color] button caused the light to turn on." Responses could range from –100 (*[color] button prevented the light from turning on*) to 100 (*[color] button caused the light to turn on*), with zero labeled *[color] button had no influence on the light turning on*. Participants also rated how confident they were in each of their causal judgments on a scale from 1 (*not at all confident*) to 7 (*very confident*).

Results and Discussion

Experiment 1A. Figure 6 shows participants' causal judgments. A one-way repeated measures analysis of variance (ANOVA) on causal judgments of unobserved causes revealed a significant effect of contingency, $F(3, 69) = 21.93$, $p < .0001$, $\eta_p^2 = 43.19$, demonstrating that the participants gave varied but systematic causal judgments of the unobserved causes. Such systematic judgments would have been highly unlikely had participants felt that there were no unique solutions for unobserved causal strengths.

Furthermore, participants' confidence ratings suggest they felt that unobserved causal strengths were not difficult to compute (Figure 7). First, confidence ratings for unobserved causes were significantly higher than the midpoint of the scale (all ps $< .05$).

[8] R-W measures association (i.e., the degree to which the cause and effect co-occur). Association, like a regression weight, does not distinguish between sufficiency and necessity. Association is simply a holistic measure of the strength of a relationship. Pearl (2000) has noted that ΔP, a quantity that R-W often mirrors at asymptote, can be interpreted as the probability of necessity and sufficiency.

Second, participants' confidence ratings for observed causes (M = 4.88, SD = 1.59) did not differ from those for unobserved causes (M = 4.83, SD = 1.51; ps > .30 in all four conditions).

What was the main factor for the systematic unobserved causal judgments in Experiment 1A? As can be seen in Figure 6, unobserved causes received much higher ratings when $\bar{O}E$ observations were included (i.e., the Unnecessary and the Zero conditions; M = 73.35, SD = 26.80) than when they were not included (i.e., the Perfect and the Insufficient conditions; M = 19.88, SD = 23.49), $t(23) = 6.57$, $p < .0001$, as predicted by BUCKLE (Figure 5). The effect of $\bar{O}E$ observations makes sense because participants were told that there were only two causes, and the effect that occurred in the absence of the observed cause could only have been brought about by the unobserved cause. According to BUCKLE, when $q_U \geq 0$ (i.e., when U is not preventative) and $O = 0$ and $E = 1$, the probability that the unobserved cause is present is always 1 (see Appendix, Equations A1 and A3). Because the unobserved cause is inferred to be present with a certainty during $\bar{O}E$ observations, the strength of the unobserved cause would increase significantly during these trials.[9]

Participants did not appear to rely on the discounting principle to estimate the strength of the unobserved causes according to the strength of the observed causes. Whereas the causal power of the observed cause (e.g., Cheng, 1997) was the same in the Unnecessary and Perfect conditions, participants judged the unobserved cause to be significantly stronger in the former than in the latter, $t(23) = 6.72$, $p < .0001$. Whereas ΔP for the observed cause was the same in the Unnecessary and Insufficient conditions, participants judged the unobserved cause to be significantly stronger in the former than in the latter, $t(23) = 5.68$, $p < .0001$.

Model performance in Experiment 1A. We compared the participants' judgments with the models described above. Recall that because Experiment 1 was designed to provide an initial look at whether people would give any systematic estimates on unobserved cause (which they did), it was not meant to offer a critical test for distinguishing the models (see the subsequent experiments). As described below, all models fit participants' data to some extent.

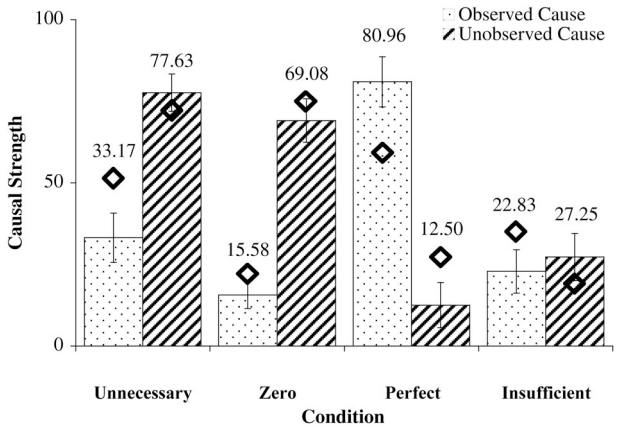

Figure 6. Causal strength judgments from Experiment 1A. Error bars indicate standard errors. Diamonds represent strength estimates made by BUCKLE (bidirectional unobserved cause learning).

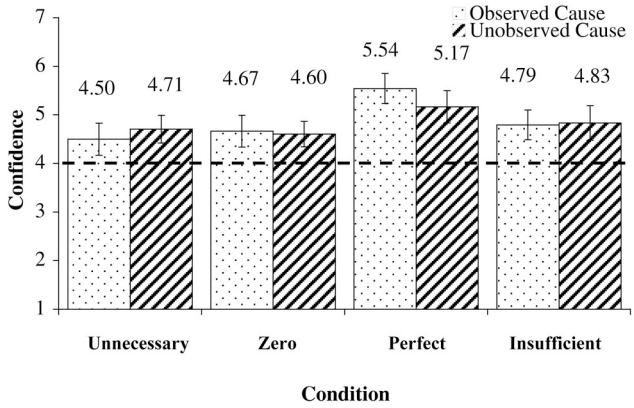

Figure 7. Confidence judgments from Experiment 1A. Error bars indicate standard errors. Dashed line indicates the midpoint of the scale.

The power models provide estimates of only the observed cause, and these estimates are illustrated in Figure 5. There was only a moderate amount of correspondence between these and participants' judgments (R^2 = .51, root-mean-squared deviation [RMSD][10] = 38.11), especially in the Unnecessary contingency, where participants' average judgment (M = 33.17) was significantly less than the predicted 100, $t(23) = 8.83$, $p < .0001$. Amending the power models with the assumption that $P(U = 1) = .5$ allows for predictions about the unobserved cause strength. With these additional predictions, the amended power models produced a better fit (R^2 = .67, RMSD = 31.95), with unobserved cause estimates diverging most greatly in the Zero condition, where participants' judgments (M = 69.08) were significantly less than 100, $t(23) = 4.66$, $p < .0001$, and in the Insufficient condition, where participants' judgments (M = 27.25) were significantly greater than 0, $t(23) = 3.75$, $p < .001$.

As discussed earlier, the power models can provide causal strength estimates of only the observed cause. To generate predictions about the unobserved causes, we assumed that $P(U = 1) = .5$. With this additional constraint, the amended power models were able to produce a fair fit (R^2 = .67, RMSD = 31.95). It should be noted that there is one particularly salient deviation from the participants' estimates: Whereas participants' average judgment for O in the Unnecessary contingency was very low (M = 33.17), the power models predicted 100, $t(23) = 8.83$, $p < .0001$. Also note that these are judgments of the observed cause where both the original and amended power models make the same prediction. Indeed, it is mainly for this reason that the fit of the original power models' predictions to the observed cause judgments is quite poor (R^2 = .51, RMSD = 38.11). It appears that participants lowered their estimates of O in the unnecessary condition because of their belief about a strong U. The power models, in their current form, are not amenable to such competition. This

[9] In BUCKLE, this is implemented by the fact that α_U would reach maximal levels (see Equation 14) and thus lead to large changes in q_U (see Equation 13).

[10] RMSD is simply the square root of the mean squared error and provides a measure of the absolute deviation, rather than the measure of relative deviation provided by R^2.

aspect is further discussed as we examine BUCKLE's performance in more detail (e.g., in Experiment 3). Next, we simulated the current experiment with the power variant of R-W (Danks et al., 2003) and BUCKLE. The observations from each contingency were presented to these models in the same order in which participants saw them. For each model, two free parameters were fitted to participants' average causal judgments.[11] The amended power variant of R-W—that is, assuming $P(U = 1) = 1$ (see introduction)—provided a reasonable fit to participants' judgments ($R^2 = .68$, RMSD = 16.59). BUCKLE, which dynamically determines the occurrence of the unobserved cause, provided the best fit ($R^2 = .79$, RMSD = 12.90, see Appendix Table A2, for fitted parameter values). Subsequent experiments compare all of these models and their underlying assumptions more critically.

Experiment 1B. Figure 8 shows participants' causal ratings from Experiment 1B. A one-way repeated measures ANOVA on causal judgments of unobserved causes revealed a significant effect of contingency, $F(3, 87) = 22.78$, $p < .0001$, $\eta_p^2 = 40.80$, showing again that participants gave varied but systematic causal judgments of the unobserved causes. Unobserved causes were again judged to be stronger when $\bar{O}E$ observations were included (i.e., the Unnecessary and the Zero conditions; $M = 69.00$, $SD = 28.55$) than when they were not (i.e., the Perfect and the Insufficient conditions; $M = 6.28$, $SD = 34.18$), $t(29) = 8.14$, $p < .0001$.

Just as in Experiment 1A, the discounting principle could not account for participants' unobserved causal judgments. Participants judged the unobserved cause to be significantly stronger in the Unnecessary condition than in the Perfect condition, $t(29) = 6.73$, $p < .0001$ (which equate the causal power of the observed cause). The unobserved cause was also judged to be significantly stronger in the Unnecessary condition than in the Insufficient condition, $t(29) = 3.43$, $p < .01$ (which equate ΔP).

Model performance in Experiment 1B. The causal strength predictions of ΔP are illustrated in Figure 5. Despite the fact that ΔP is unable to produce estimates for the unobserved causes, there was a good fit between its estimates and participants' judgments of the observed causes ($R^2 = .85$, RMSD = 17.26). Turning to R-W, we again fitted its two free parameters to participants' mean judgments using the method described above, treating the constantly present experimental context as though it were equivalent to the unobserved cause of interest. R-W was able to fit the data well ($R^2 = .77$, RMSD = 21.97). Again, subsequent experiments compare these models more critically.

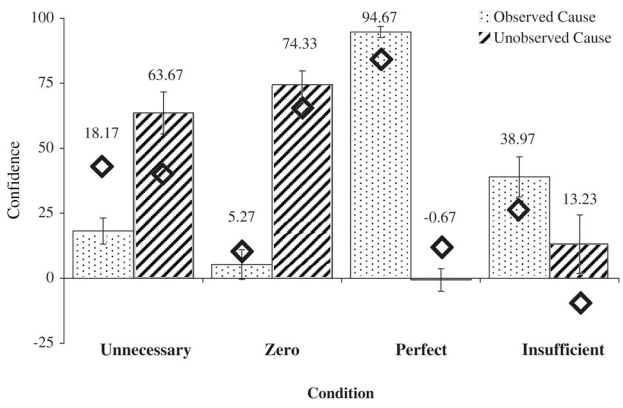

Figure 8. Causal strength judgments from Experiment 1B. Error bars indicate standard errors. Diamonds represent the Rescorla–Wagner model's strength estimates.

Summary. Experiment 1 provides an initial look at unobserved cause learning. Given the overt lack of information about part of the causal system, learners could have provided random judgments of the unobserved cause. Yet participants appeared relatively unfazed by the obvious lack of information and provided systematic responses about the unobserved cause. These findings allow us to rule out the models in the first row of Table 1, which predict that participants should be agnostic about strength of the unobserved cause.

One methodological limitation of the current experiment is that the task may have encouraged participants to learn about the unobserved causes, because they were told ahead of time that they would be asked to evaluate the strength of the unobserved causes. Participants were also reminded about the existence and possible operation of the unobserved causes on every trial (see Figure 2), which may have led them to dedicate unnatural attention to the unobserved causes. Experiment 2 examines this possibility.

Experiment 2: Learning Without Explicit Instructions About Unobserved Causes

In Experiment 2, we removed all mention of unobserved causes during learning. Only after the learning portion of the experiment had concluded were participants told about the possibility of an unobserved cause. This way, the learning phase of the experiment was no different from that of a traditional causal learning experiment, concerning only the learning of observed causes.

Method

Experiment 2 was a 2 (Explicit vs. Implicit) × 2 (Unnecessary vs. Insufficient; see Figure 5) between-subjects design. The Explicit condition ($n = 24$) proceeded just as in Experiment 1. Participants in the Implicit condition ($n = 26$) were told that the system included one button and a light and were not given any initial information about the existence (or nonexistence) of alternative causes. Only after viewing the entire set of trials were these participants told that the initial description of the system was incomplete and that they would be asked to judge a second button whose information had been lost. All participants then evaluated each button as in Experiment 1A.

Results and Discussion

The pattern of results in the Implicit condition (Figure 9) was identical to that found in Experiments 1A and 1B. The unobserved cause was judged to be significantly stronger in the Unnecessary

[11] For the power variant of R-W, the two free parameters were α_1 and α_0, the learning rate parameters associated with a present and absent cause, respectively. For BUCKLE, the two free parameters were α_O and α_U, the learning rate parameters associated with the observed and unobserved cause, respectively. The best fitting values of these parameters were found using the directed search algorithm described by Hooke and Jeeves (1960).

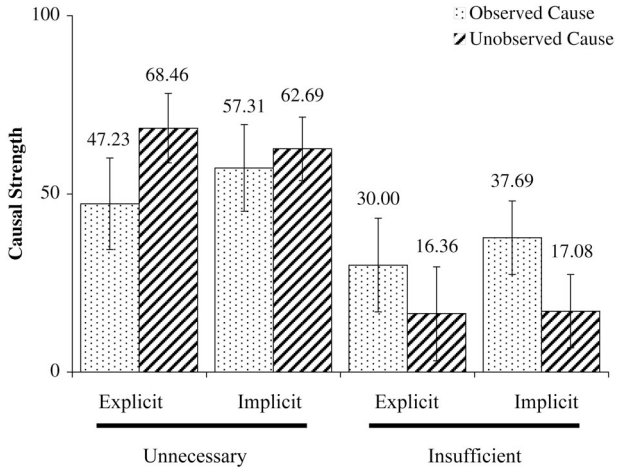

Figure 9. Causal strength judgments from Experiment 2. Error bars indicate standard errors.

than in the Insufficient condition even when participants had no information about an alternative cause until they were asked to make judgments.[12] A 2 (Explicit vs. Implicit) × 2 (Unnecessary vs. Insufficient) ANOVA on the ratings of the unobserved cause showed significantly higher ratings in the Unnecessary condition ($M = 65.58$, $SD = 40.57$) than in the Insufficient condition ($M = 16.75$, $SD = 31.53$), $F(1, 46) = 21.28$, $p < .0001$. Of note, both the main effect of the instructional manipulation (i.e., Explicit vs. Implicit) and the interaction effect were nonsignificant (all Fs < 0.90).

Summary of Experiments 1 and 2

The results of Experiments 1 and 2 allow us to begin evaluating the models under consideration. Models that ignore unobserved causes or are otherwise unable to produce unobserved cause judgments (the models in the top row of Table 1) will be unable to account for participants' willingness (Luhmann & Ahn, 2003) and ability to provide systematic unobserved cause judgments. Thus, our subsequent investigation is dedicated to those models that are able to produce unobserved cause estimates. Experiment 1 found that the remaining models exhibited varying degrees of success, with BUCKLE exhibiting a slightly better fit than the rest. However, because it is difficult to interpret small differences in fit, we turn our attention to the larger conceptual differences between the remaining models. The first of these differences we consider relates to how the unobserved cause occurs: Is it constant, or does it vary from trial to trial (i.e., the middle vs. bottom row in Table 1)?

Experiment 3: Evaluating Beliefs About the Occurrence of Unobserved Causes

The main goal of Experiment 3 was to explore whether and how people's judgments about the occurrence of the unobserved cause vary. On every trial, participants were presented with information about the presence or absence of one of the causes and the effect, while the second cause remained unobserved, and were asked to rate the probability that the unobserved cause was present on each trial.

As explained in the introduction, the most obvious innovation of BUCKLE is its dynamic inferences about the occurrence of the unobserved cause. In contrast, the amended power and R-W models (the middle row of Table 1) are unable to dynamically infer the occurrence of an unobserved cause on the basis of the type of trial, or prior learning, and are able to estimate the causal strength of the unobserved cause only because they assume that the unobserved cause is constantly present as part of the constantly present background. Thus, the first question is whether people's judgments about the occurrence of the unobserved cause vary or remain static.

The second question is, if the unobserved cause is believed to vary, how does it vary? BUCKLE's operation suggests that even for unobserved causes, causal strength estimates and covariation are intimately related. For example, BUCKLE predicts that the stronger one believes the unobserved, generative cause is, the more one would believe that U would covary with E (and vice versa). These predictions are derived from Equations A1 and A2 in the Appendix. As the strength of the unobserved cause (q_U) goes to 1, $P(U = 1 \mid O = o, E = 1)$ increases (it approaches $\frac{1}{[(o \cdot q_O) + 1]}$) and $P(U = 1 \mid O = o, E = 0)$ decreases (it approaches 0). Thus, as the strength of the unobserved cause increases, it is more likely to be present when the effect is present and less likely to be present when the effect is absent. Experiment 3 examines these detailed predictions of BUCKLE.

Method

The method was the same as in Experiment 1 (Figure 5) except that after each trial was presented, participants ($N = 24$) judged, "How likely is it that the [button corresponding to the unobserved cause] was pressed in this test?" (0 = *definitely NOT pressed*; 10 = *definitely pressed*). Asking participants about the probability of the unobserved cause on every trial did not appear to have created an unusually disruptive learning situation, because the pattern of causal strength judgments (see Figure 10) mirrored that of Experiment 1.[13] The remaining procedure and the design were the same as in Experiment 1A except that the cell counts of 7 in Figure 5 were changed to 10 to increase the number of measurements.

[12] One might argue that participants made judgments about the unobserved cause retrospectively when they were asked about the unobserved cause, rather than spontaneously making inferences about the unobserved cause during learning. Experiment 4 shows that this was unlikely to be the case.

[13] A one-way repeated measures ANOVA on causal judgments of unobserved causes revealed a significant main effect of contingency, $F(3, 66) = 10.37$, $p < .0001$ (one participant failed to provide judgments for the perfect condition). As in Experiment 1, the effect of contingency is in large part due to participants giving much higher ratings on conditions with $\bar{O}E$ ($M = 72.60$, $SD = 23.73$) than on conditions without $\bar{O}E$ ($M = 41.47$, $SD = 23.08$).

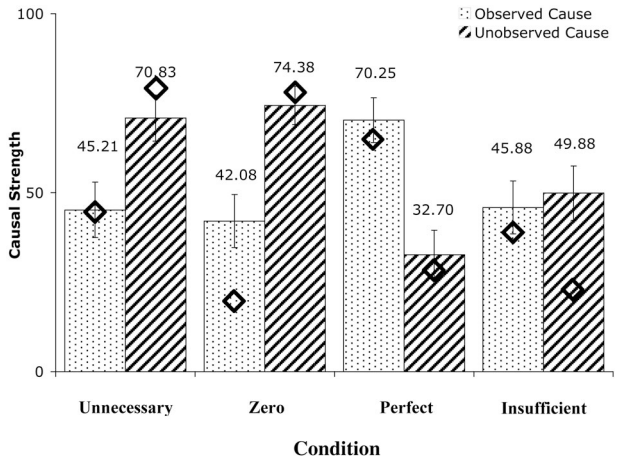

Figure 10. Causal strength judgments from Experiment 3. Error bars indicate standard errors. Diamonds represent estimates made by BUCKLE (bidirectional unobserved cause learning).

Results and Discussion

Figure 11 shows the mean probability judgments broken down by contingency and trial type.[14] First, note that participants believed that the unobserved cause was neither constantly present (i.e., the assumption allowing estimation of unobserved causal strengths from the R-W models; see introduction) nor equally likely on all trials (i.e., the assumption allowing unobserved causal strengths from the amended power models). Instead, probability judgments varied considerably and systematically, as BUCKLE assumes. For example, participants' probability judgments varied as a function of the type of observation (e.g., $O\bar{E}$, OE). The main effect of trial type from one-way repeated measures ANOVAs was significant in three conditions (all ps $< .05$) and marginal in the perfect condition, $F(1, 23) = 4.08$, $p = .055$.

Participants' probability judgments also suggest that they expected the unobserved cause to covary with the effect, as implemented in BUCKLE. As shown by the marginal averages below each matrix in Figure 11, participants believed the unobserved cause to be more likely present when the effect was present than when the effect was absent. This finding makes sense given that participants' causal strength judgments for the unobserved cause were greater than zero in all four conditions; positive covariation led to positive causal judgments, just as in observed cause learning.

Taking this a step further, participants may have believed the unobserved cause covaried with the effect more in those situations where the unobserved cause was judged to be strong than in those situations where the unobserved cause was judged to be weak. To explore this possibility, we compared probability judgments during OE trials and $\overline{OE}$ trials, because these trial types were shared across the four contingencies, and therefore, any differences found cannot have been due to the trial type being judged. If participants believed the unobserved cause varied with the effect, they should believe the unobserved cause to be more likely present on OE trials and less likely present on $\overline{OE}$ trials.

We subtracted the average rating for $\overline{OE}$ trials from the average rating for OE trials. This composite score served as an index of the degree to which participants believed the unobserved cause to covary with the effect on these trials. We then correlated this index with the group's mean causal judgments of the unobserved cause within each condition. The correlation was significant, $r(4) = .98$, $p < .05$, suggesting that beliefs about stronger covariation (i.e., higher composites) were associated with beliefs about a stronger unobserved cause. Here again we find that the process of unobserved cause learning shares many of the characteristics of observed cause learning.

We were also interested in the degree to which participants believed the unobserved cause varied with the observed cause. As mentioned in the introduction, for the power models to operate correctly, unobserved causes must be independent of observed causes—that is, $P(O \mid U = 1) = P(O \mid U = 0)$ (see Cheng, 1997). In contrast, recent work (Hagmayer & Waldmann, 2007) suggests that participants do not necessarily share this assumption. To examine this issue in the current study, we averaged probability ratings for trials in which the observed cause was present—that is, $P(U \mid O)$—and for trials in which the observed cause was absent—that is, $P(U \mid \sim O)$—separately for each participant and each condition.[15] The difference between these quantities was significant in the Unnecessary condition (mean difference = .97), $t(23) = 2.18$, $p < .05$; the Zero condition (mean difference = –.43), $t(23) = 2.12$, $p < .05$; and the Insufficient condition (mean difference = 1.80), $t(23) = 4.38$, $p < .001$, and was marginally significant in the Perfect condition (mean difference = .94), $t(23) = 2.02$, $p = .055$. These findings mirror those of Hagmayer and Waldmann (2007) and suggest that learners are not making the assumptions required by the power models.

BUCKLE's performance. Because participants' probability estimates were variable (a finding that cannot be accounted for by any of the power or R-W models), we simulated Experiment 3 using only BUCKLE. The best fitting parameters allowed BUCKLE to account for 76% of the variance in participants' causal judgments (RMSD = 13.65). These same parameter values were then used to derive probability estimates of U for each trial in each contingency. These values (multiplied by 10 to match the scale used by participants) can be seen in Figure 11. We note several important features of the probability values.

First, BUCKLE predicts that probability estimates should differ depending on trial type (e.g., OE vs. $\overline{OE}$). To quantitatively evaluate this effect, we averaged probability estimates separately for each trial type in each condition for both participants and BUCKLE. The probability estimates generated by BUCKLE accounted for a significant amount of variance in participants' probability judgments ($R^2 = .91$, RMSD = 1.60).

[14] The results reported here concern only the probability judgments averaged across the trial positions because participants' probability judgments hardly varied as a function of trial position. BUCKLE's probability estimates are similarly flat. See Luhmann (2006, Experiment 5) for more detailed results.

[15] By averaging across different trial types (e.g., $\overline{OE}$ and $\bar{E}$) to derive an estimate of the conditional probability—for example, $P(U \mid O = 0)$—we must assume that our participants' probability judgments were mapped onto the response scale in a more or less linear manner. Violations of this assumption could distort further interpretation. Nonetheless, because this is the methodology used by Hagmayer and Waldmann (2007, Experiment 1), we are able still to compare our findings with theirs.

Condition

	Unnecessary			Zero			Perfect			Insufficient		
		E	$\overline{E}$		E	$\overline{E}$		E	$\overline{E}$		E	$\overline{E}$
Participants	O	5.80		5.80	O	5.38	**3.41**	4.40	O	4.10		4.10	O	5.25	**3.76**	4.51
	$\overline{O}$	**7.51**	2.15	4.83	$\overline{O}$	**7.84**	1.82	4.83	$\overline{O}$		3.16	3.16	$\overline{O}$		2.7	2.7
		6.66	2.15			6.61	2.62			4.10	3.16			5.25	3.23	
BUCKLE		E	$\overline{E}$		E	$\overline{E}$		E	$\overline{E}$		E	$\overline{E}$	
	O	7.30		7.30	O	7.91	**3.13**	5.52	O	5.56		5.56	O	5.71	**4.59**	5.15
	$\overline{O}$	**10**	3.06	6.53	$\overline{O}$	**10**	3.22	6.61	$\overline{O}$		4.50	4.50	$\overline{O}$		4.60	4.60
		8.65	3.06			8.95	3.17			5.56	4.50			5.71	4.59	

Figure 11. Average trial-by-trial probability judgments of participants and predictions of BUCKLE for the various trial types in each condition of Experiment 3. BUCKLE = bidirectional unobserved cause learning.

Second, as can be seen in Figure 11, the probability of U being present is much higher when the effect is present (e.g., OE and $\bar{O}E$) than when the effect is absent (e.g., OE or $\overline{O\bar{E}}$), illustrating that BUCKLE, like the participants, predicts covariation between U and E. Composite scores for each contingency (i.e., difference between probability estimates from $\overline{O\bar{E}}$ and OE trials) using BUCKLE (4.25, 4.69, 1.07, and 1.11 for the Unnecessary, Zero, Perfect, and Insufficient conditions, respectively) were highly correlated with BUCKLE's own strength estimates for the unobserved cause, $r(4) = .99$, $p < .05$, just as they were for participants' judgments.

Summary. The results of Experiment 3 indicate that learners do not believe that the unobserved, alternative cause is constantly present or present with a fixed probability across all trial types and conditions. This suggests that the assumptions needed to amend the power and R-W models to obtain predictions about unobserved causal strengths do not accurately describe the manner in which people learn about unobserved causes. Instead, learners appear to make sophisticated inferences about the occurrence of unobserved, alternative causes. Judgments about the occurrence of the unobserved cause varied greatly as a function of whether the observed cause and the effect were present. Judgments also varied systematically, even for identical observations. Contingencies that elicited strong causal judgments of the unobserved cause led participants to believe that the unobserved cause varied with the effect more than conditions that elicited weaker unobserved cause judgments. These findings suggest that beliefs about causal strength (e.g., the perceived strength of the unobserved cause) and beliefs about the occurrence of the unobserved cause are intimately related, just as for observed causes.

With these results we may now conclude that the most accurate explanation for people's unobserved cause learning occupies the bottom row of Table 1. As discussed in the introduction, EM cannot produce unique estimates of unobserved causal strength, and thus we are left with BUCKLE. As discussed above, we believe that BUCKLE's iterative process is critical to its ability to successfully perform this task. With Experiment 4 we attempt to provide more direct evidence that people's behavior reflects a specifically iterative process.

Experiment 4: Order Effects

In Experiment 4, we manipulated the order in which specific types of observations were encountered and examined the resulting change in causal strength judgments. The set of trials used is illustrated in the top panel in Figure 12. This set of trials was divided into two blocks. One of the blocks was analogous to the Unnecessary condition in Experiment 1 and contained $\bar{O}E$, $\bar{O}\bar{E}$, and OE observations. The other block was analogous to the Insufficient condition in Experiment 1 and contained $O\bar{E}$, $\bar{O}\bar{E}$, and OE observations. Participants received these blocks in one of two orders: Unnecessary followed by Insufficient (Unnecessary–Insufficient condition) or Insufficient followed by Unnecessary (Insufficient–Unnecessary condition; see the bottom panel

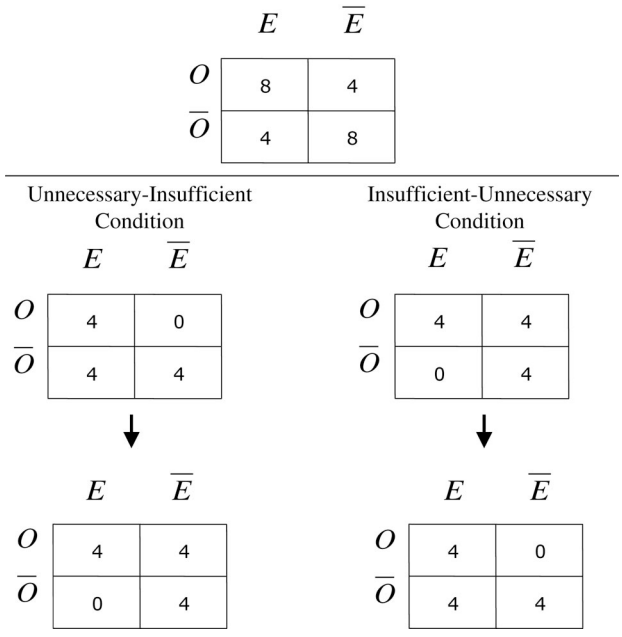

Figure 12. Illustration of the design used in Experiment 4. The two sets of contingency tables in the bottom panel show the order in which participants received two blocks of trials in each condition. Each ordering results in the identical contingency, as described by the table at the top.

in Figure 12). Note that because the only manipulation was the order of the two blocks, participants always saw the same set of observations by the end of the sequence.

The simultaneous models (i.e., the left column of Table 1) are, by definition, unable to account for trial order effects, because their computations are performed over the entire set of observations at once. Thus, any order effects exhibited by participants' judgments would support an iterative model such as BUCKLE.

BUCKLE specifically predicts that unobserved cause judgments should differ between the two orderings. To see why this is, consider the Unnecessary–Insufficient order. During the first block of this condition, $\bar{O}E$ observations will lead to the unobserved cause being perceived as strong (as illustrated in Experiment 1). When the second block (without $\bar{O}E$ observations) is encountered, the now-strong unobserved cause will be interpreted as covarying with the effect (as illustrated in Experiment 3). For instance, a learner would believe that the unobserved cause would likely be present during OE trials but absent in $O\bar{E}$ and $\bar{O}\bar{E}$ trials. These inferences would further increase the strength of the unobserved cause.

Now consider the Insufficient–Unnecessary condition, in which $\bar{O}E$ observations are encountered in the second half. In this situation, at the end of the first half, the unobserved cause will be perceived as weak (as illustrated in the Insufficient condition of Experiment 1). Only when the second block (with $\bar{O}E$ observations) is encountered will the perceived strength of the unobserved cause begin to increase. Thus, only the second block in this order will increase the strength of the unobserved cause. As compared with the Unnecessary–Insufficient order, there are far fewer observations that act to increase the strength of the unobserved cause. Thus, the unobserved cause should be perceived as stronger when $\bar{O}E$ observations are encountered in the first block than when they are encountered in the second block.

Though Experiment 3 did not provide support for R-W and the power variant of R-W, Experiment 4 provides another opportunity to test these models in a more natural context than Experiment 3. It is interesting to note that these models make qualitatively different predictions from BUCKLE in Experiment 4, because they are unable to modulate $P(U = 1)$ on the basis of prior learning. Consider the Unnecessary–Insufficient order. Like BUCKLE, these models predict that $\bar{O}E$ observations act to increase the strength of the unobserved cause during the first block (e.g., see R-W's predictions for the unnecessary condition in Figure 8). However, whereas this strength leads BUCKLE to reduce the probability of the unobserved cause occurring during the second block's $\bar{O}\bar{E}$ and $O\bar{E}$ observations, both R-W models assume that the unobserved cause is definitely present during both $\bar{O}\bar{E}$ and $O\bar{E}$ observations. The unexpected absence of an effect in the presence of the unobserved cause during these observations (i.e., large "error") then leads to a significant drop in the strength of the unobserved cause. Thus, the Unnecessary–Insufficient order leads both the R-W and power variant of R-W to predict relatively weak judgments of the unobserved cause.

In contrast, consider the Insufficient–Unnecessary order. Like BUCKLE, the R-W models predict that the unobserved cause gains only small amounts of causal strength during the first block (e.g., see R-W's prediction for the insufficient condition in Figure 8). However, whereas this weakness leads BUCKLE to be relatively agnostic about the occurrence of the unobserved cause during the second block's OE observations, the R-W models continue to assume that the unobserved cause is definitely present during OE observations. The unexpected presence of an effect in the presence of the unobserved cause during these observations then leads to a significant increase in the strength of the unobserved cause. Thus, the Insufficient–Unnecessary order leads both the R-W and the power variant of R-W to predict relatively strong judgments of the unobserved cause.

To summarize, BUCKLE and the R-W models differ in their predictions about the influence of the order manipulation. BUCKLE uses the beliefs developed in the first half of the sequence to modulate beliefs about the occurrence of the unobserved cause in the second half. R-W and the power variant of R-W cannot adjust the occurrence of the unobserved cause on the basis of prior learning and thus predict that the second half of the sequence is processed in a relatively unbiased manner. Thus, Experiment 4 provides another test of the difference between the R-W models' constantly present unobserved cause and BUCKLE's dynamically occurring unobserved cause.

Method

The stimulus materials were similar to those used in Experiment 1. The cell frequencies of each condition are summarized in Figure 12. In the Unnecessary–Insufficient condition, participants first saw the block containing $\bar{O}E$ trials followed by the block containing $O\bar{E}$ trials. In the Insufficient–Unnecessary condition, participants saw the two blocks in the reverse order. Although the sequence of trials was made of two blocks, there was nothing noting the change from one block to the other, and as far as participants were concerned, they were experiencing one continuous stream of observations. The procedure of Experiment 4 was the same as in Experiment 1A. Each participant saw both orders instantiated in different colored buttons, and the orders were counterbalanced across participants ($N = 50$).

Results and Discussion

As summarized in Figure 13, participants gave a significantly higher rating for the unobserved cause in the Unnecessary–Insufficient condition ($M = 73.50$, $SD = 28.79$) than in the Insufficient–Unnecessary condition ($M = 61.66$, $SD = 34.81$), $t(49) = 2.89$, $p < .01$, even though they embodied identical contin-

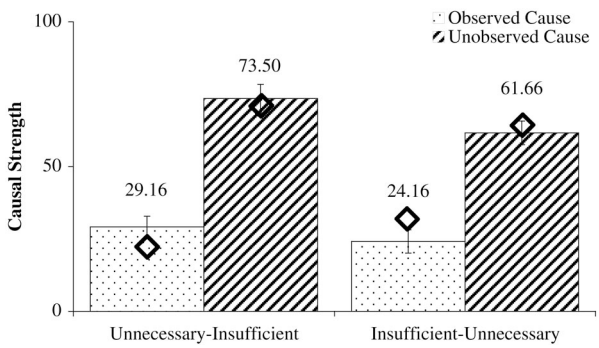

Figure 13. Causal strength judgments from Experiment 4. Error bars indicate standard errors. The diamonds represent estimates made by BUCKLE (bidirectional unobserved cause learning).

gencies overall. These results clearly indicate that simultaneous models will be unlikely to account for people's unobserved cause learning.

To test iterative models, we simulated BUCKLE as well as R-W and the power variant of R-W (Danks et al., 2003) for each of the conditions used in Experiment 4 by presenting the models with the exact same set of observations in the exact same order in which participants received them, fitting the free parameters as before. BUCKLE predicted the unobserved cause to be stronger in the Unnecessary–Insufficient condition ($q_U = 69.19$) than in the Insufficient–Unnecessary condition ($q_U = 64.28$; see Figure 13). These estimates accounted for 91% of the variance in participants' judgments for both unobserved and observed causes (RMSD = 5.79). Both the traditional and power variants of R-W predicted an order effect but did so in the wrong direction. These models estimated the unobserved cause to be stronger in the Insufficient–Unnecessary condition (R-W: $q_U = .56$; power variant: $q_U = .51$) than in the Unnecessary–Insufficient condition (R-W: $q_U = .32$; power variant: $q_U = .36$), resulting in poor fit ($R^2 = .52$, RMSD = 19.22 for the traditional R-W; $R^2 = .57$, RMSD = 17.53 for the power variant of R-W).

BUCKLE's unique success in the current experiment suggests that the observed findings result from dynamic beliefs about the occurrence of the unobserved cause. Experiment 4 also provides additional evidence for the details of BUCKLE's unobserved cause learning process. Experiment 3 suggested that participants were able to provide systematic causal strength estimates of unobserved causes even when they had no a priori knowledge of the existence of alternative causes. Experiment 4 suggests that people make spontaneous inferences about the occurrence of unobserved causes despite never being prompted to do so. These results also illustrate that inferences about the occurrence of unobserved causes can have a real impact on learning and, ultimately, causal strength judgments.

Experiment 5: The Influence of $O\bar{E}$ Observations

Experiment 5 investigated one finding that appears to contradict BUCKLE's behavior. Schulz and Sommerville (2006) demonstrated that preschoolers believe that $O\bar{E}$ observations suggest the preventative influence of an unobserved cause. The current study, however, has failed to obtain similar effects. For example, the insufficient condition in Experiment 1A included $O\bar{E}$ observations, but the unobserved cause was judged to be generative (e.g., $M = 27.25$).

This apparent discrepancy can be explained as follows. According to BUCKLE, $O\bar{E}$ observations can occur for two separate reasons. For example, if a learner believes, on the basis of previous trials, that O is generative and U is preventative, then an $O\bar{E}$ observation may occur (a) because U prevented the effect from happening or (b) because O is not entirely sufficient to bring about the effect (see Equation A6 in the Appendix). In the latter case, it is ambiguous whether U is present or absent, and thus, the amount of learning taking place about U during $O\bar{E}$ observations would be reduced, resulting in weak (but not necessarily preventative) causal strength for U, as observed in the current study.

BUCKLE predicts, however, that if the observed cause were sufficiently strong (as preschoolers might have believed in Schulz & Sommerville's [2006] study), then $O\bar{E}$ observations would more strongly indicate the influence of a preventative unobserved cause. More specifically, the term $[1-o \cdot q_O]$, from the denominator of Equation A6 in the Appendix, will decrease, increasing $P(U = 1)$. With the unobserved cause now likely to be present in the absence of the effect, q_U will decrease (i.e., a stronger preventative cause). Experiment 5 tested these predictions by training participants, before observing $O\bar{E}$, to believe the observed cause was strong.

Method

Each participant was randomly assigned to either the Unnecessary condition ($n = 22$) or the Insufficient condition ($n = 22$). The trial sequence was divided into two phases. The first phase was the same for both conditions and was designed to increase the perceived strength of the observed cause without changing beliefs about the unobserved cause. During the first phase (20 trials), both of the causes were observable. The cause that was to be unobserved in the second phase was explicitly noted to be absent on every trial (see Figure 14), whereas the other cause varied across trials with ΔP of .8. The second phase (12 trials) was similar to previous experiments; the unobserved cause was unobserved, and the observed cause remained observed. In the Insufficient condition, the second phase included $O\bar{E}$, OE, and $\bar{O}\bar{E}$ observations, and in the Unnecessary condition it included $\bar{O}E$, OE, and $\bar{O}\bar{E}$ observations. At the end of the second phase, participants rated the causal strength of the observed and unobserved cause as in previous experiments.

New stimulus materials were developed, because the buttons and lights used as stimuli previously are more compatible with generative than with preventative causal relations (i.e., buttons normally do not prevent lights from being turned on). Stimuli in Experiment 5 were novel medications (e.g., "DJE-143") and physical side effects (e.g., "salivation increased/did not increase").

Results and Discussion

As shown in Figure 15, the unobserved cause was judged to be significantly preventative in the insufficient condition ($M = -12.27$, $SD = 23.51$), one-sample test against zero: $t(21) = 2.45$, $p < .05$, replicating Schulz and Sommerville (2006), whereas in the unnecessary condition it was judged to be significantly positive ($M = 34.77$, $SD = 23.25$), one-sample test against zero: $t(21) = 7.01$, $p < .0001$. Furthermore, judgments of the observed cause in the insufficient condition were significantly lower than those in the unnecessary

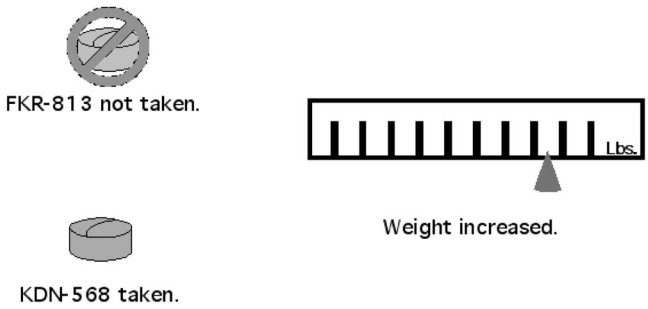

Figure 14. A sample trial used in the first phase of Experiment 5. One of the causes (the bottom one) is observed throughout the entire experiment. The other cause (the top one) is observed and constantly absent in the first phase. In the second phase, this cause will become unobserved, just as in previous experiments.

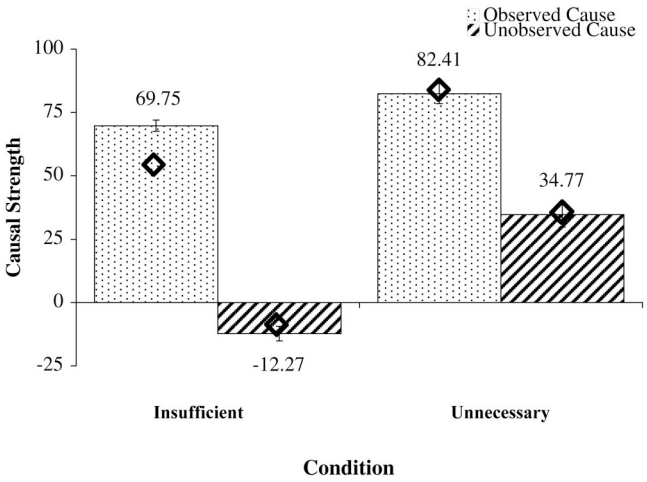

Figure 15. Causal strength judgments from Experiment 5. Error bars indicate standard errors. Diamonds represent estimates made by BUCKLE (bidirectional unobserved cause learning).

condition, $t(42) = 6.67$, $p < .0001$, indicating that $O\bar{E}$ observations decreased the perceived strength of the observed cause.

BUCKLE accounted for 95% of the variance (RMSD = 9.04). Unlike all previous simulations, BUCKLE's causal strength estimate of the unobserved cause was preventative (i.e., negative) in the Insufficient condition ($q_U = -5.07$), whereas it remained positive in the Unnecessary condition ($q_U = 35.75$; see Figure 15).

It should also be noted that, like participants' judgments, BUCKLE's estimate of the unobserved cause was only weakly preventative in the Insufficient condition. According to BUCKLE (see Appendix), even when q_U is small (e.g., –.01) and q_O is large (e.g., 0.9), the probability of U being present will be only slightly greater than chance (e.g., .52). It is only when the perceived strength of the observed cause is near maximal that $O\bar{E}$ observations will strongly implicate the presence of the unobserved cause (e.g., if $q_O = .999$ and $q_U = -.01$, then $P(U = 1) = .92$). Thus, BUCKLE makes the intuitive suggestion that the strength of the observed cause will predict the magnitude of the influence of $O\bar{E}$ observations on unobserved cause judgments. If, as Schulz and Sommerville (2006) suggested, their preschoolers believed the observed cause to be absolutely sufficient, they naturally would have inferred the influence of a strong preventative unobserved cause.

To summarize, BUCKLE argues that in Experiments 1–3, $O\bar{E}$ observations did not lead learners to perceive the unobserved cause as preventative, because $O\bar{E}$ observations could have been due to weak q_O or negative q_U. When pretraining in Experiment 5 established O to be a strong generative cause, participants inferred a preventative unobserved cause from $O\bar{E}$ observations. BUCKLE also explains why $O\bar{E}$ observations led to only weakly preventative causal judgments of the unobserved cause. Only when the observed cause is nearly absolutely sufficient (perhaps as in the beliefs of preschoolers; Schulz & Sommerville, 2006) will the unobserved cause be perceived as strongly preventative.

General Discussion

The data reported above suggest that people are able to learn about unobserved causes. Current models of causal learning appear unable to account for this behavior even when amended with additional assumptions. In contrast, the model proposed here, BUCKLE, appears capable of explaining a variety of unobserved cause learning phenomena. BUCKLE embodies a simple two-step process for learning in the presence of unobserved causes. The first step is to compute how likely the unobserved cause is to be present. This inference is made using information available in the environment (e.g., the state of observed causes and effects), as well as empirically derived beliefs (e.g., beliefs about the strengths of the causes). After this inference is made, there is no longer any missing information; all causes are believed to be present with some probability. The second step is to learn about the underlying causal relationships. To do this, BUCKLE makes a prediction about how likely the effect is to be present given the currently available information (e.g., the state of observed and unobserved causes and effects and beliefs about the causes' strengths). This inference allows BUCKLE to compare its beliefs about how the world operates (i.e., the probability inference) with the actual operation of the world (i.e., whether the effect actually occurred). Errors in this step's inference then form the basis of learning. Despite its relative simplicity, BUCKLE appears to accurately capture a significant variety of aspects of people's cause learning.

Our first main finding is that people provided systematic estimates about strengths of unobserved causes, a finding that several prominent models of causal induction (e.g., ΔP) cannot explain. Second, as predicted by BUCKLE, people's estimates of the likelihood that an unobserved cause is present on each trial were varied and sensitive to both the current observation and beliefs about the causal strengths held during that trial. Third, people's unobserved cause learning was sensitive to trial order, a finding that can be explained only by iterative models such as BUCKLE. BUCKLE was also able to account for the fact that participants did not necessarily judge the unobserved cause to be preventative in the presence of $O\bar{E}$ observations, a finding that appears to contradict previously reported findings in children (Schulz & Sommerville, 2006). However, we were able to use BUCKLE to infer what the underlying developmental difference might be and design a situation that elicited preventative judgments, providing further evidence in support of BUCKLE's account and reconciling BUCKLE with the previous findings.

BUCKLE's Assumptions

BUCKLE's performance, as described throughout the present article, is the result of an array of assumptions, some of which are made to represent specific causal systems used in the experiments and others concerning how causal inferences would proceed. Here, we briefly outline the assumptions underlying the current simulations to illustrate that representational assumptions are interchangeable whereas processing assumptions are meant to be psychological claims.

Representational assumptions. We have assumed a particular parameterization: Causes combine in the manner of a noisy-OR/noisy-AND-NOT. This assumption has received much recent support (e.g., Cheng, 1997; Griffiths & Tenenbaum, 2005) and is intuitively appropriate given the stimuli we have used here. Given other situations, however, other parameterizations would be more appropriate (e.g., Waldmann, in press). For example, the overall temperature of a room might be the summed result of multiple sources of heat (e.g., sunlight, heater, body temperature), as R-W

assumes. In such cases, we would expect that people's behavior would change. BUCKLE can easily be changed to reflect the new parameterization. The result would be a two-step algorithm in which the first step replaces the missing presence/absence data (the equations in the Appendix would change) and the second step is equivalent to R-W's learning rule.

In addition, we have performed the current simulations under the simplified assumption that there are no alternative causes (other than the single unobserved cause). As mentioned above, we are not committed to this assumption as a psychological claim and instead suggest that it would be highly influenced by experimental instruction, domain, and background knowledge. Again, only a simple modification of BUCKLE would be required to make changes in that assumption. Nevertheless, it is interesting to note that when our participants were not explicitly told to make this assumption (Experiment 2), the pattern of judgments did not change. Further empirical work will be required to determine what shapes learners' beliefs about background causes and what their default assumptions are.

Another interchangeable, representational assumption made in the current simulation of BUCKLE is about specifics of how the occurrence of the unobserved cause is inferred. First, when utilizing Bayes's theorem to compute the probability that the unobserved cause is present on each trial (BUCKLE's Step 1), we used a uniform prior of $P(U = 1 \mid O = 1) = P(U = 1 \mid O = 0)$. This is a simplistic assumption, but it still allowed BUCKLE to account for more than 90% of the variance in people's probability judgments. Here again, we expect that the experimental methodology and background knowledge will likely influence the prior required to successfully model people's behavior, but uniform priors seem to be reasonable defaults.

Second, we have assumed that this prior does not change over the course of learning. This is likely a more surprising assumption because the posteriors computed on previous trials should intuitively influence the prior on subsequent trials. However, this seems not to be necessary, at least in simulating the current experiments. There are two pieces of evidence for this claim. First, in Experiment 3, participants' trial-by-trial probability judgments for a given trial type showed very little change over time (see Luhmann, 2006, for details). If the prior were to change with experience, one would expect these judgments (i.e., the posteriors) to change as well (though how much change to expect could vary). Second, we have simulated the entire set of experiments using a version of BUCKLE in which the prior was updated during learning. These modifications caused BUCKLE to deviate from both participants' causal strength and trial-by-trial probability estimates. Furthermore, Hagmayer and Waldmann (2007) reported similar results. Their participants were asked to provide judgments of either the prior—for example, $P(U = 1 \mid O = o)$—or the posterior—for example, $P(U = 1 \mid O = o, E = e)$. Whereas judgments of the posterior varied depending on the input, judgments of the prior "hardly showed any systematic relation to the data" (Hagmayer & Waldmann, 2007, p. 351). Taken together, these findings suggest that fixed priors may be a useful way of characterizing our learners' behavior.

Processing assumptions. Much more critical to the current findings are BUCKLE's assumptions about processing. These assumptions also served as a framework for distinguishing the existing models of causal learning, as outlined in Table 1.

First, BUCKLE was conceived as an iterative learning model. This aspect of BUCKLE's process appears to be necessary to explain the current results (e.g., the order effect in Experiment 4), though one could imagine a similar model operating over large amounts of data at once (see below for examples).

Second, as we saw throughout the evaluation, the models that provided or assumed some value of $P(U = 1)$ (e.g., the amended power and R-W models and BUCKLE) were able to provide causal strength estimates of unobserved causes, whereas those that did not estimate $P(U = 1)$ could not (e.g., power models, ΔP). Given that people were willing to provide systematic estimates of the causal strength of unobserved causes, providing some estimate of $P(U = 1)$ appears to be more psychologically valid.

Third, assuming a fixed value of $P(U = 1)$ is not sufficient for accurately mirroring participants' behavior. Instead, BUCKLE's dynamic probability estimates, influenced by values of O and E as well as q_U and q_O, were able to provide better quantitative and qualitative fits to our participants' behavior than those provided by static estimates of $P(U = 1)$.

Unlike the representational assumptions outlined above, these last three processing assumptions are indispensable to BUCKLE's operation. Removing or drastically altering these portions of BUCKLE would result in a qualitatively different model. However, the current experiments provide significant evidence to support these aspects of BUCKLE as psychologically valid.

BUCKLE and the EM Algorithm

As mentioned in the introduction, the process described by BUCKLE has substantial similarities with the EM algorithm (Dempster et al., 1977). The EM algorithm is designed to accomplish learning despite incomplete data. EM then estimates the values of both the missing data and the parameters. The essential component of EM's operation is an alternation between two steps. The first step is to compute a likelihood distribution over all possible values of the missing data based on the current estimates of the parameters. This step essentially acts to "fill in" any missing data. The second step is to adjust the value of the parameters being learned (e.g., the strengths of the causes) on the basis of the inference-amended data that were generated in the first step. Alternating between these two steps has been shown to converge on locally maximal parameter estimates (Dempster et al., 1977, provided the proof).

The operation of the EM algorithm is obviously quite similar to that of BUCKLE. Both algorithms perform two steps: one to replace missing data and one to update the parameters of interest (q_O and q_U). Both models perform each step using the results from the previous step, essentially ignoring the fact that there is missing data (and that the results from the previous steps are only approximate). Unlike BUCKLE, the original EM algorithm was designed to perform its two-step procedure over an entire data set at once. However, researchers have developed versions of EM that operate incrementally (e.g., Bradley, Fayyad, & Reina, 1998; Suematsu, Maebashi, & Hayashi, 2004; see also Neal & Hinton, 1998). The only difference between BUCKLE and a full-blown incremental EM algorithm is that BUCKLE only considers the current observation when updating the causal strength estimates. However, the historical roots of BUCKLE's trial-by-trial strength updating (e.g., Rescorla & Wagner, 1972) strongly suggest that such updating is psychologically plausible. Future theoretical work will be required to evaluate the consequences of these differences.

Finally, it should be noted that BUCKLE is not the only example of EM to be applied to cognitive processes. Fried and Holyoak (1984) put forth a model (the category density model) of unsupervised category learning that is isomorphic to EM. The missing data in their study were the exemplar-category assignments (i.e., the traditional feedback in category learning experiments), and the estimated parameters were the category prototypes (plus category variability). The success of BUCKLE and the category density model suggests that the EM algorithm may characterize a cognitive strategy for dealing with missing data across many domains (e.g., category learning, causal learning).

References

Anderson, J. R., & Sheu, C. (1995). Causal inferences as perceptual judgments. *Memory & Cognition, 23,* 510–524.

Bradley, P. S., Fayyad, F. M., & Reina, C. A. (1998, November). *Scaling EM (expectation–maximization) clustering to large databases* (Report MSR-TR-98–35). Redmond, WA: Microsoft.

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

Busemeyer, J. R. (1991). Intuitive statistical estimation. In N. H. Anderson (Ed.), *Contributions to information integration theory* (Vol. 1, pp. 187–215). Hillsdale, NJ: Erlbaum.

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

Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive causal power. In F. Keil & R. Wilson (Eds.), *Cognition and Explanation* (pp. 227–253). Cambridge, MA: MIT Press.

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

Cheng, P. W., Park, J., Yarlas, A. S., & Holyoak, K. J. (1996). A causal-power theory of focal sets. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 313–357). San Diego, CA: Academic Press.

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

Danks, D., Griffiths, T. L., & Tenenbaum, J. B. (2003). Dynamical causal learning. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* (pp. 67–74). Cambridge, MA: MIT Press.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–38.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Goodie, A. S., Williams, C. C., & Crooks, C. L. (2003). Controlling for causally relevant third variables. *Journal of General Psychology, 130,* 415–430.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 354–384.

Hagmayer, Y., & Waldmann, M. R. (2007). Inferences about unobserved causes in human contingency learning. *Quarterly Journal of Experimental Psychology, 60,* 330–355.

Hooke, R., & Jeeves, T. A. (1960). "Direct search" solution of numerical and statistical problems. *Journal of the Association for Computing Machinery, 8,* 212–229.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79,* 1–17.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones et al. (Eds.), *Attribution: perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.

Luhmann, C. C. (2005). Confounded: Causal inference and the requirement of independence. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1355–1360). Mahwah, NJ: Erlbaum.

Luhmann, C. C. (2006). *BUCKLE: A model of causal learning.* Unpublished doctoral dissertation, Vanderbilt University, Nashville, Tennessee.

Luhmann, C. C., & Ahn, W. (2003). Evaluating the causal role of unobserved variables. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 734–739). Mahwah, NJ: Erlbaum.

Luhmann, C. C., & Ahn, W. (2005). The meaning and computation of causal power: A critique of Cheng (1997) and Novick and Cheng (2004). *Psychological Review, 112,* 685–692.

Neal, R. M., & Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan (Ed.), *Learning in graphical models* (pp. 355–368). Cambridge, MA: MIT Press.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111,* 455–485.

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

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

Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development, 77,* 427–442.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120.

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

Shanks, D. R. (1989). Selectional processes in causality judgments. *Memory & Cognition, 17,* 27–34.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48,* 257–279.

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

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B, 13,* 238–241.

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7,* 337–342.

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

Suematsu, N., Maebashi, K., & Hayashi, A. (2004). An online variant of EM algorithm based on the hierarchical mixture model learning. In M. H. Hamza (Ed.), *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications* (pp. 270–275). Anaheim, CA: ACTA Press.

Waldmann, M. R. (in press). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science.*

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

White, P. A. (2002). Causal attribution from covariation information: The evidential evaluation model. *European Journal of Social Psychology, 32,* 667–684.

(*Appendix follows*)

Appendix

Probability Computations for BUCKLE

Table A1 includes an explanation of the notation used throughout the article along with their initial values (e.g., at the beginning of an experiment) and how they change during the course of learning. Table A2 includes the computations that allow BUCKLE to replace the information missing from the input. These expressions are simply Bayes' theorem (below) expanded assuming the noisy-OR/noisy-AND-NOT parameterization (see the main text for details). Table A3 includes the best fitting parameter values in each of the reported experiments.

$$P(U = 1|O = o,E = e) = \frac{P(E = e|O = o,U = 1) \cdot P(U = 1|O = o)}{[P(E = e|O = o,U = 0) + P(E = e|O = o,U = 1)]\cdot P(U = 1|O = o)}$$

Assuming that $P(U = 1|O = 0) = P(U = 1|O = 1)$:

$$P(U = 1|O = o,E = e) = \frac{P(E = e|O = o,U = 1)}{P(E = e|O = o,U = 0) + P(E = e|O = o,U = 1)}.$$

Table A1

Parameters Used by BUCKLE Along With Descriptions of Their Nature and Use

Name	Initial value	Source	Description
O	N/A	Input	Presence/absence of the observed cause on the current observation
E	N/A	Input	Presence/absence of the effect on the current observation
q_O	0	Learned	Causal sufficiency of the observed cause
q_U	0	Learned	Causal sufficiency of the unobserved cause
$P(U\|O = 1)$	.5	Static	Prior likelihood of the unobserved cause being present during O observations
$P(U\|O = 0)$	.5	Static	Prior likelihood of the unobserved cause being present during $\bar{O}$ observations
α_O	N/A	Fit to data	Learning rate associated with the observed cause
α_U	N/A	Fit to data	Learning rate associated with the unobserved cause
β	.5	Static	Learning rate associated with the effect

Table A2

Expressions Used to Replace Missing Information About Unobserved Causes

Situation	$P(U = 1\|O = o, E = 1)$	$P(U = 1\|O = o, E = 0)$
$q_o \geq 0$, $q_u \geq 0$	$\frac{(o \cdot q_O) + q_U - (o \cdot q_O \cdot q_U)}{(o \cdot q_O) + [(o \cdot q_O) + q_U - (o \cdot q_O \cdot q_U)]}$	$\frac{1 - [(o \cdot q_O) + q_U - (o \cdot q_O \cdot q_U)]}{[1 - (o \cdot q_O)] + \{1 - [(o \cdot q_O) + q_U - (o \cdot q_O \cdot q_U)]\}}$
$q_o < 0$, $q_u \geq 0$	$\frac{q_U \cdot [1 - (o \cdot q_O)]}{0 + \{q_U \cdot [1 - (o \cdot q_O)]\}} = 1$	$\frac{1 - \{q_U \cdot [1 - (o \cdot q_O)]\}}{1 + [1 - \{q_U \cdot [1 - (o \cdot q_O)]\}]}$
$q_o \geq 0$, $q_u < 0$	$\frac{o \cdot q_O \cdot (1 - q_U)}{(o \cdot q_O) + [o \cdot q_O \cdot (1 - q_U)]}$	$\frac{1 - [o \cdot q_O \cdot (1 - q_U)]}{[1 - (o \cdot q_O)] + \{1 - [o \cdot q_O \cdot (1 - q_U)]\}}$
$q_o < 0$, $q_u < 0$	$\frac{0}{0 + 0}$ = undefined	$\frac{1}{1 + 1} = .5$

Table A3

Best-Fitting Values of α_O and α_U in Each Experiment

Experiment	α_O	α_U
1A	0.269	0.229
1B	0.280	0.199
3	0.218	0.176
4	0.122	0.222
5	0.273	0.194
M (*SD*)	0.234 (0.066)	0.204 (0.022)

Received December 31, 2005
Revision received February 7, 2007
Accepted February 7, 2007 ■

Correction to Nelson (2005)

In the article "Finding Useful Questions: On Bayesian Diagnosticity, Probability, Impact, and Information Gain," by Jonathan D. Nelson (*Psychological Review*, 2005, Vol. 112, No. 4, pp. 979–999), there was a typographical error in the data for percentage of females with short and long hair in Table 13 on p. 992. The data should indicate that 7% of females had short hair and 93% of females had long hair. The calculations and discussion in the article were based on these correct percentages.

DOI: 10.1037/0033-295X.114.3.677