Modeling the Role of Unobserved Causes in Causal Learning

Christian C. Luhmann (christian.luhmann@vanderbilt.edu)
Department of Psychology, 2 Hillhouse Ave
New Haven, CT 06511 USA

Woo-koung Ahn (woo-kyoung.ahn@yale.edu)
Department of Psychology, 2 Hillhouse Ave
New Haven, CT 06511 USA

Abstract

Current theories suggest that causal learning is based on covariation information. However, information about the presence/absence of events (particularly causes) is frequently unavailable, rendering them unobserved. The current paper presents a new model of causal learning, BUCKLE (Bidirectional Unobserved Cause LEarning), which extends existing models of causal learning by dynamically inferring information about unobserved causes. During the course of causal learning, BUCKLE continually computes the probability that an unobserved cause is present on each occasion and uses the results of these inferences to adjust the strengths of the unobserved, as well as observed, causes.

Keywords: Causal learning, inference, induction

Introduction

Current models of causal induction assume that the input available to reasoners comes in the form of covariation; how the causes vary with their effects. Thus, a learner observes whether the presence or absence of a causal candidate is followed by the presence or absence of an effect, and translates these observations into beliefs about causal relations.

Yet, in the real world, covariation is often not available. For example, acquiring information about the presence/absence of causes sometimes requires special methods (e.g., genetic influences on cancer). Perhaps more commonly, causes are unobserved simply because learners cannot possibly consider all alternative causes of a particular event. For instance, we do not know all possible causes for gender discrepancy in science. Thus, lacking information about the presence/absence of causes seems to be the rule rather than the exception. This paper presents a new model of causal learning, BUCKLE, which attempts to capture how people learn causal relations when information about causes is missing.

BUCKLE

BUCKLE (Bidirectional Unobserved Cause LEarning) assumes that the learning environment always includes an unobserved cause and learns by performing two steps during each trial. The first step is to compute the probability that the unobserved cause is present. The second step is to adjust the strengths of each cause-effect relationship using an error-correction algorithm.

To compute the probability of the unobserved cause (u) in a situation with one observed cause (o) and one effect (e), BUCKLE applies Bayes theorem to the current beliefs about the strength of o and u (q_o and q_u respectively) and the prior belief about the probability of u being present (i.e., P(u) with we will always assume to be .5). The following equations are for cases when q_o and q_u are believed to be generative in a current trial (see Luhmann & Ahn, 2006, for equations for other cases):

$$P(u \mid o=0, e=0) = \frac{P(u) \bullet (1 - q_u)}{\left[1 - P(u)\right] + \left[P(u) \bullet (1 - q_u)\right]} \quad (1)$$

$$P(u \mid o=0, e=1) = \frac{P(u) \bullet q_u}{P(u) \bullet q_u} = 1 \quad (2)$$

$$P(u \mid o=1, e=0) = \frac{(1-q_o) \bullet P(u) \bullet (1 - q_u)}{\left\{(1-q_o) \bullet \left[1 - P(u)\right]\right\} + \left[(1-q_o) \bullet P(u) \bullet (1 - q_u)\right]} \quad (3)$$

$$P(u \mid o=1, e=1) = \frac{P(u) \bullet \left[q_o + q_u - (q_o \bullet q_u)\right]}{\left\{q_o \bullet \left[1 - P(u)\right]\right\} + \left\{P(u) \bullet \left[q_o + q_u - (q_o \bullet q_u)\right]\right\}} \quad (4)$$

In words, Equation 1, for instance, shows the probability that u is present when o and e are absent. The denominator is the probability of e being present given that o is absent, which occurs when either u is absent, or u is present but fails to cause e. The numerator of Equation 1 is the probability of the latter occurring (i.e., u being present). Once the probability of u is computed, the unobserved cause is treated just like an observed cause except that it is present with some probability.

These equations allow BUCKLE to make several predictions. For example, the probability of u should vary as a function of trial type (i.e., whether o and e are present or absent). Also, note that Equation 2 is special. This equation suggests that people should believe that u present with a certainty (i.e., $P(u) \approx 1$), when o is absent but e is present (i.e., what Luhmann & Ahn, 2003 call *unexplained effects*), because o and u are the only possible causes.

BUCKLE's second step is to use the observed and inferred information to adjust the strength of each causal relationship. BUCKLE learns via an error-correction algorithm. Information about the state of the causes (i.e., o and u) is first used to predict how likely the effect is given BUCKLE's current causal beliefs (i.e., q_o and q_o). This

prediction is then compared with the actual presence/absence of the effect. The difference between the predicted and actual states of the effect (the prediction error) forms the basis of learning. BUCKLE predicts the effect according to equation 5:

$$e_{predicted} = P(e) = (o \bullet q_o) + (u \bullet q_u) - \left[(o \bullet q_o) \bullet (u \bullet q_u)\right] \quad (5)$$

In this expression, o=1 or 0 (when the observed cause is present or absent, respectively) and u=P(u | o, e). This expression (as well as Equations 1-4) assumes that causes combine in the manner of a noisy-OR gate (e.g., Cheng, 1997; Danks, Tenenbaum, & Griffiths, 2003; Griffiths & Tenenbaum, 2005).

Based on the prediction error, the strength of each cause is updated separately:

$$\Delta q_o = \alpha_o \beta (e - e_{predicted}) \quad (6)$$

$$\Delta q_u = \alpha_u \beta (e - e_{predicted}) \quad (7)$$

The quantities α and β represent learning rates associated with the causes and effects, respectively. A value of 0.5 will be used for β. When the observed cause is present, α_o=$\alpha_{o\text{-}present}$ where $\alpha_{o\text{-}present}$ will be treated as a free parameter and allowed to vary between zero and one. When the observed cause is absent, α_o=α_{absent}=0.0. For the unobserved cause, Equation 8 is used to compute a value of α_u to take into account the fact that the unobserved cause is only present with some probability.

$$\alpha_u = \left[P(u) \bullet (\alpha_{u-present} - \alpha_{absent})\right] + \alpha_{absent} \quad (8)$$

When P(u)=0, this equation results in α_u=0; when P(u)=1, α_u=$\alpha_{u\text{-}present}$, just as for the observed cause. For values of P(u) between 0 and 1, α_u increases linearly and in proportion to the value of P(u). The variable $\alpha_{u\text{-}present}$ will be treated as a second free parameter and allowed to vary between zero and one.

BUCKLE makes several novel predictions about the causal strength of unobserved causes. For example, as explained above, unexplained effects should lead to the belief that u is present with a certainty in the presence of e. Thus, unexplained effects should act to greatly increase q_u (because α_u will also be maximal, see Equation 8). Indeed, Luhmann and Ahn (2003) demonstrated that unobserved cause judgments were heavily influenced by the occurrence of unexplained effects.

BUCKLE also makes predictions about the inferred probability of u. For example, probability judgments should vary systematically depending on the presence/absence of o and u, as illustrated in Equations 1-4. Furthermore, judgments about the presnce/absence of u should be related to q_u. For example, BUCKLE predicts that positive values of q_u should be accompanied by beliefs about positive covariation between u and e (i.e., P(u|e=1) - P(u|e=0) > 0). Additionally, beliefs about the occurrence of the unobserved cause should be correlated with subsequent causal strength judgments. These predictions will be further illustrated by using BUCKLE to simulate the experiments below.

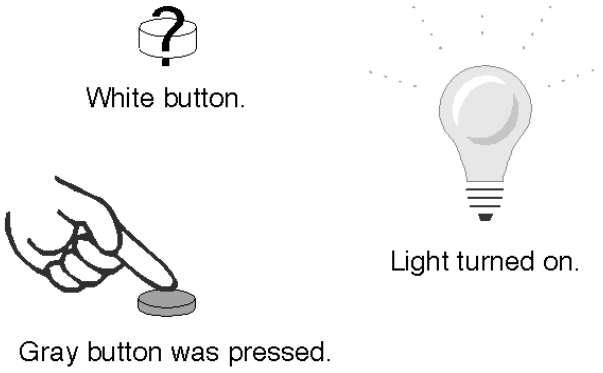

Figure 1 – Example stimuli. The unobserved cause is denoted by the large "?".

Experiment 1

To test BUCKLE's predictions, we used a learning setting with one effect, one observed cause, and one unobserved cause whose state (present vs. absent) was unknown to participants (see Figure 1). Experiment 1 examines how well BUCKLE accounts for (1) people's causal strength judgments, (2) people's probability judgments of the unobserved cause, and (3) the relationship between these two judgments.

Method

Twenty-four Vanderbilt University undergraduates participated in Experiment 1. Stimuli consisted of novel electrical systems. Each system contained one button whose state (pressed or not) was observable, one button whose state was unobservable and a single light. The unobserved button was marked with a large question mark to denote the lack of presence/absence information (see Figure 1). Participants were told that it was their job to determine how the systems worked and that they would be asked to judge the extent to which each button caused the light to turn on.

Figure 2 illustrates the contingency between the observed cause and the effect for each of the four conditions. The Zero condition contained both $\bar{O}E$ and $O\bar{E}$ observations with the contingency between o and e being zero (ΔP=0). The Perfect condition contained neither $\bar{O}E$ nor $O\bar{E}$ observations (ΔP=1). The remaining two conditions each constituted moderately strong relationships (ΔP=0.5). The

Condition	Unncessary		Zero		Perfect		Insufficient	
	E	$\bar{E}$	E	$\bar{E}$	E	$\bar{E}$	E	$\bar{E}$
Contigency Structure	O 10	0	O 10	**10**	O 10	0	O 10	**10**
	$\bar{O}$ **10**	10	$\bar{O}$ **10**	10	$\bar{O}$ 0	10	$\bar{O}$ 0	10

Figure 2 – The four contingencies used in Experiment 1 and 2.

Unnecessary condition included $\overline{O}E$ observations (i.e., unexplained effects), which render the observed cause partially unnecessary but completely sufficient. The Insufficient condition included $O\overline{E}$ observations, which render the observed cause partially insufficient but completely necessary.

Each participant saw all four conditions separately in a counterbalanced order. For a given condition, participants received the included observations in a pseudo-random order. On each trial, participants were presented with information about the presence/absence of the observed cause and the effect (e.g., Figure 1). After receiving this information, participants were immediately asked to judge how likely the unobserved cause was to be present on that occasion. This judgment was made on a scale of 1 ("Not at all likely") to 9 ("Definitely likely"). Once this judgment was made, the next trial began. After all the observations in the condition were presented, participants were asked to judge the causal strength of the observed and unobserved causes.

Results

Causal Strength Judgments. Figure 3 shows participants' mean causal strength judgments. To examine how causal strength judgments varied across the four contingencies, we performed a 2 ($\overline{O}E$ present/absent) X 2 ($O\overline{E}$ present/absent) repeated measures ANOVA on causal judgments of the unobserved causes. This analysis revealed a significant main effect of $\overline{O}E$ information, $F(1, 22) = 26.59$, $p<.0001$, because participants gave much higher ratings on conditions with $\overline{O}E$ observations (M=72.60, SD=28.77) than on conditions without $\overline{O}E$ observations (M=41.47, SD=35.69). No other main effects or interactions were significant. Note that these results imply that the strength of the unobserved cause is not simply inversely proportional to that of the observed cause (e.g., in the Insufficient condition) as one might expect given an account that emphasizes discounting (e.g., Thagard, 2000). These results also closely mirror those of Luhmann and Ahn (2003) who found that observations of $\overline{O}E$ exerted a particularly strong influence on causal strength judgments.

We applied BUCKLE to the exact same set of observations in the exact same order that participants received them. BUCKLE's final causal strength estimates accounted for 81% of the variance in participants' causal judgments (RMSD of 13.25). Importantly, BUCKLE accounts for the large influence of $\overline{O}E$ observations on judgments of the unobserved cause.

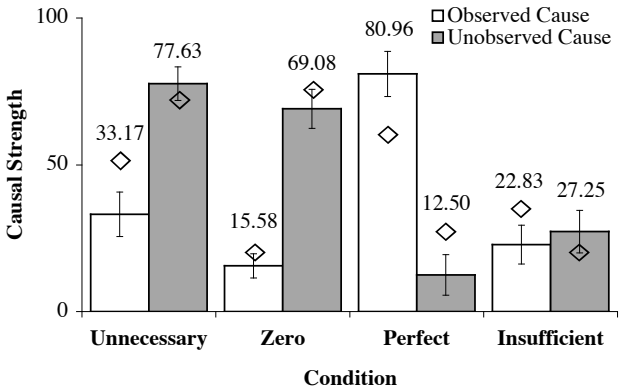

Figure 3 – Causal strength judgments from Experiment 1. Error bars illustrate standard error and the diamonds illustrate BUCKLE's estimates.

Probability Judgments. Figure 4 shows, broken down by condition and trial type, participants' mean probability judgments of the likelihood that *u* is present in a trial. Individual one-way repeated measures ANOVAs were performed on each of the four conditions with trial type as the independent factor. The effect of trial type was significant in three of the four conditions (all p's < .05) and marginally significant in the Perfect condition ($F(1,23) = 3.67$, $p=.068$). Thus, as predicted by BUCKLE, participants appear to be making varied, but systematic inferences about the presence of the unobserved cause (cf. Rescorla & Wagner, 1972). Also, note that the unobserved cause was judged to be most likely present during $\overline{O}E$ observations (unexplained effects) as predicted by BUCKLE's (see Equation 2).

To quantitatively evaluate the fit between participants' estimates and BUCKLE's predictions, we compared participants' average probability judgments for each trial type (e.g., $\overline{O}E$, $O\overline{E}$) in each condition with BUCKLE's estimates. BUCKLE's estimates provided a good fit, accounting for a significant amount of variance in participants' judgments ($R^2=.86$, RSMD=1.48). These results (collapsed across condition) are shown in Figure 5.

It is also interesting to note that participants' probability judgments imply a positive correlation between the presence of the unobserved cause and the presence of the effect. This can be seen by looking at the marginal averages below each matrix in Table 2; the unobserved cause was judged to be more likely to occur when *e* was present than when *e* was absent. This finding makes sense given that participants' causal strength judgments of the unobserved cause were greater than zero in all four conditions. Current theories of

Condition

Likelihood(*u*)

Unncessary	E	$\overline{E}$	
O	5.8		5.8
$\overline{O}$	**7.51**	2.15	4.83
	6.66	2.15	

Zero	E	$\overline{E}$	
O	5.38	**3.41**	4.4
$\overline{O}$	**7.84**	1.82	4.83
	6.61	2.62	

Perfect	E	$\overline{E}$	
O	4.1		4.1
$\overline{O}$		3.16	3.16
	4.1	3.16	

Insufficient	E	$\overline{E}$	
O	5.25	**3.76**	4.51
$\overline{O}$		2.7	2.7
	5.25	3.23	

Figure 4 – Probability judgments from Experiment 1. Marginal averages below each matrix illustrate that participants' believe the unobserved cause to vary with the effect.

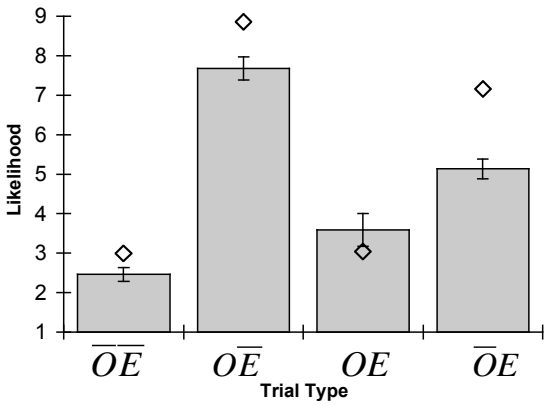

Figure 5 – Likelihood judgments for each trial type collapsed across contingency. Error bars illustrate standard error and the diamonds illustrate BUCKLE's predictions (again collapsed over contingency)

causal learning, including BUCKLE, imply that positive covariation should accompany positive causal judgments.

Taking this idea a step further, there should have been a strong relationship between participants' beliefs about the occurrence and strength of the unobserved cause. To evaluate this prediction, we compared participants' probability judgments from OE trials and $\overline{O}\overline{E}$ trials (these were the only trial types shared across the four conditions). If participants believed the unobserved cause varied with the effect, they should have judged the unobserved cause to be more likely present on OE trials and less likely on $\overline{O}\overline{E}$ trials. If participants did not believe that the unobserved cause covaried with the effect, they should have believed that the probability of the unobserved cause was more similar on these two trial types.

Each participant's average probability judgment for $\overline{O}\overline{E}$ trials was subtracted from their average probability judgment for OE trials separately for each condition. This composite score served as a measure of the degree to which participants believed the unobserved cause to covary with the effect on these trials. Note that the composite for each condition was computed using identical observations. Nonetheless, the composite accounted for nearly all the variance in participants' average causal strength judgments (R^2=.96).

To test whether BUCKLE mirrored these beliefs, we computed a composite score (as before) using BUCKLE's probability estimates during $\overline{O}\overline{E}$ and OE trials. Just as for participants' judgments, BUCKLE's composite scores accounted for 99% of the variance in BUCKLE's final unobserved cause strength estimates.

Summary

The results of Experiment 1 illustrate several important points. First, participants were able to provide systematic causal judgments of causes that were not observed. Our own model, BUCKLE, suggests that these judgments result from a sophisticated learning process that replaces the missing information inferentially. Thus, the second finding was that, as predicted by BUCKLE, learners make dynamic inferences about the occurrence of unobserved causes. Judgments about the probability of the unobserved cause varied as a function of whether the observed cause and the effect were present. The third finding was that probability judgments varied, even during identical observations, across the different contingencies and did so systematically. Causal strength judgments of the unobserved causes were accompanied by predictable judgments about how the unobserved cause occurred in the presence and absence of the effect.

These findings suggest that people's beliefs about the occurrence of the unobserved cause are intimately related to the strength of that cause. Note that this is exactly what happens with observed causes. The perceived strength of an observed cause is intimately related to its presence/absence. The difference in the current situation is that participants must infer the presence/absence of the cause on their own. The fact that learning otherwise continues as normal is a testament to the resilience of the responsible processes.

BUCKLE accounts for the relationship between probability and strength judgments and suggests that the probability judgments being made on a trial-by-trial basis provide the basis for learning and subsequent causal strength judgments. Thus, BUCKLE argues that the perceived strength of the unobserved cause cannot be separated from beliefs about the way in which the unobserved cause occurs. Experiment 2 further explores this claim.

Experiment 2

One critical aspect of the learning process described by BUCKLE is that causal strength estimates are updated in a sequential manner as each observation is made. This differs from approaches that compute causal strength over all available data once enough observations have been accumulated (e.g., Cheng, 1997; White, 2002). An interesting consequence of this is that the order in which observations are encountered should influence the final causal strength estimates. This is because the probability of u being present depends on q_u and q_o, which, according to BUCKLE, change over time. Altered probability judgments might then lead to altered causal strength judgments as we saw in Experiment 1.

To test this possibility, we used the set of trials summarized in Figure 6. This set of trials was divided into two blocks. One of the blocks contained unexplained effects (analogous to the Unnecessary condition) and the other did not (analogous to the Insufficient condition). These two blocks could be ordered in one of two ways; the block containing unexplained effects could be presented either first (early-unexplained-effects condition) or second (late-unexplained-effects condition) as shown in Figure 6. Note that, because the only manipulation was the order of the two blocks, participants had always seen the same set of

observations by the end of the sequence. Thus, any differences between orders cannot be a result of the number or type of trials.

BUCKLE predicts that the judged strength of the unobserved cause will differ between the two orderings. Consider the early-unexplained-effects condition. During the first block of this condition, the unexplained effects will lead to the unobserved cause being perceived as strong (as illustrated in Experiment 1). When the second block (without unexplained effects) is encountered, the strong unobserved cause will be interpreted as covarying strongly with the effect (also illustrated in Experiment 1). For instance, a learner would believe that the unobserved cause would likely be present during OE trials but likely absent during $O\overline{E}$ trials. These inferences should lead to further increases in the strength of the unobserved cause.

In contrast, consider the late-unexplained-effects condition in which the unexplained effects are encountered at the end. In this situation, at the end of the first half, the unobserved cause will be perceived as weak (as illustrated in Experiment 1). Only once the unexplained effects in the second block are encountered will the perceived strength of the unobserved cause will begin to increase. However, compared to the early-unexplained-effects condition, there are far fewer trials acting to increase the perceived strength of the unobserved cause. Thus, the unobserved cause will be perceived as stronger when encountering unexplained effects in the first block than when encountering them in the second block.

Method

Fifty undergraduates from Vanderbilt University participated for partial fulfillment of course credit. The stimulus materials were similar to Experiment 1. The statistical properties of the system are summarized by the cell frequencies illustrated in Figure 6.

The sole manipulation in this experiment was the order in which trials were presented to participants. There were two orderings used, each of which consisted of two blocks. One block contained $O\overline{E}$ trials but not $\overline{O}E$ trials. The other contained $\overline{O}E$ trials but not $O\overline{E}$ trials. In the early-unexplained-effects condition, participants first saw the block containing $\overline{O}E$ trials followed by the block containing $O\overline{E}$ trials. In the late-unexplained-effects condition, participants saw the two blocks in the reverse order. Although there were two blocks, there was nothing noting the change from one block to the other, and as far as participants were concerned, they were experiencing a continuous stream of observations.

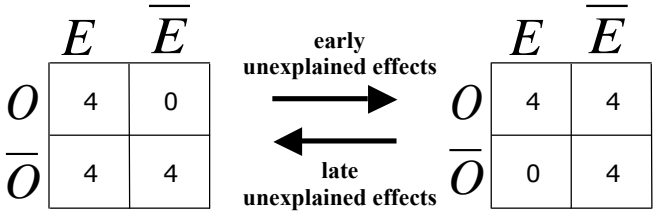

Figure 6 – The design of Experiment 2. Two blocks of trials were presented in two orders.

The procedure of Experiment 2 was the same as Experiment 1 except that probability judgments were not elicited. After completing observations, participants were asked to judge the causal strength of each cause. Each subject saw both orders instantiated with different color buttons with the order of the two sequences counterbalanced across participants.

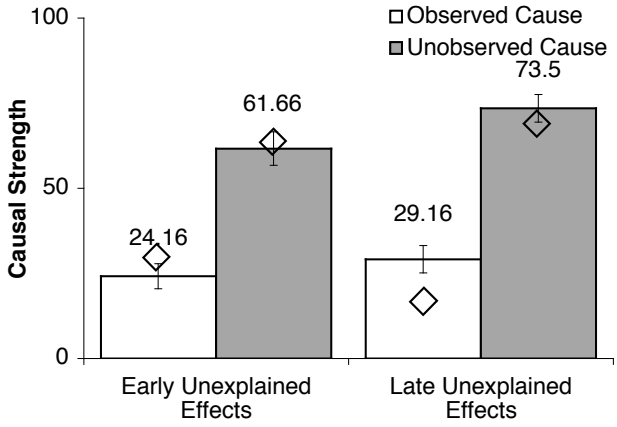

Figure 7 – Causal strength judgments from Experiment 2. Error bars illustrate standard error and the diamonds illustrate BUCKLE's estimates.

Results

As Figure 7 illustrates, despite identical sets of observations, the unobserved cause was judged to be significantly stronger in the early-unexplained-effects condition (M = 73.50, SD = 25.90) than in the late-unexplained-effects condition (M = 61.66, SD = 27.79), $t(49)=2.89$, $p < .01$. Using the exact same set of observations in the exact same order that participants received them, BUCKLE's estimate of the unobserved cause's strength was higher in the early-unexplained-effects condition (q_u=69.19) than in the late-unexplained-effects condition (q_u=64.28).

Discussion

The model proposed here, BUCKLE, learns about unobserved causes using two steps. First, BUCKLE infers the probability of the unobserved cause using its current beliefs. Second, BUCKLE adjusts its beliefs about the strength of causal relationships via error correction. Despite its relative simplicity, BUCKLE appears to accurately capture a significant variety of aspects of people's causal learning.

First, BUCKLE's estimates of the causal strength of the unobserved cause mirrored those of participants. Second, Experiment 1 demonstrated that BUCKLE's estimates of the probability of the unobserved cause matched participants' own judgments. Currently, BUCKLE is the only model in the field that can make such predictions. For instance, the model proposed by Rescorla and Wagner (1972) also acknowledges the existence of an unobserved cause.

However, because this cause is treated as a part of an unchanging context, the Rescorla-Wagner model has no way of accounting for dynamic changes in the probability of an unobserved cause.

Third, and perhaps more interesting, was the relationship we observed between participants' judgments of the occurrence of the unobserved cause and their subsequent strength judgments of the unobserved cause. This finding reaffirms the idea that causal judgments are based on covariation. What is novel about this finding is that participants were not given any covariation information about the unobserved cause. The covariation between the unobserved cause and effect had to be generated by the participants themselves.

Fourth, BUCKLE accounted for the order effect found in Experiment 2. Such order effects pose problems to all models that provide causal strength estimates only at the end of learning (e.g., Cheng, 1997, White, 1992).

One potential criticism is that the current results may have been obtained simply because participants were constantly reminded of a possibility of unobserved cause during learning (i.e., Figure 1). However, Luhmann and Ahn (2003, Experiment 1) found that even when participants were explicitly allowed to refuse judgment, they were still willing to provide causal strength estimates for unobserved causes. These results suggest that the inferences about unobserved cause occur spontaneously and naturally. They also indicate the importance of further investigating the role of inferences on unobserved cause in explaining human causal learning.

Acknowledgments

This project was supported by a National Institute of Mental Health Grant (RO1 MH57737) to the Woo-kyoung Ahn.

References

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

Dennis, M. J., & Ahn, W. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory and Cognition, 29*, 152-164.

Goodie, A. S., Williams, C. C., Crooks, C. L. (2003). Controlling for causally relevant third variables. *Journal of General Psychology, 130*, 415-30.

Hagmayer, Y., & Waldmann, M. R. (in press). Seeing the unobservable - Inferring the probability and impact of hidden causes, *Quarterly Journal of Experimental Psychology*.

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* (734-739). Mahwah, New Jersey: Lawrence Earlbaum Associates, Inc.

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, New Jersey: Lawrence Erlbaum Associates, Inc.

Luhmann, C. C., & Ahn, W. (2006). *BUCKLE: A Model of Causal Learning*. Manuscript submitted for publication.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York, NY: 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*. New York, NY: Appleton-Century-Crofts.

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

Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge, Massachusetts: MIT Press.