Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 280

17

Dynamic Interpretations of Covariation Data

Woo-kyoung Ahn, Jessecae K. Marsh, & Christian C. Luhmann

In discovering causes of events, people evidently use various types of evidence or cues (e.g., Einhorn & Hogarth, 1986). Virtually all models of causal learning (e.g., Cheng, 1997; Rescorla & Wagner, 1972) have focused on how causal relations are learned based on covariation information—namely, information about whether the presence or absence of one event (C or ~C)\edq1\ co-occurs with the presence or absence of another event (E or ~E). Thus, in all of these models, relevant input data are classified as CE, ~CE, C~E, or ~C ~E, as summarized in Figure 17-1. Existing models of causal learning have stipulated different ways in which these four types of covariation evidence would or should be combined to evaluate the causal relationship among events. Yet, these models in their current forms share an underlying assumption that all events of a given type (e.g., CE) play an identical role in assessing causal strength.

One model of causal induction combines covariation information into a contingency measure called ΔP\edq2\ (e.g., Jenkins & Ward, 1965). The value of ΔP is calculated as follow:

$$\Delta P = \left(\frac{(CE)}{(CE)+(C\sim E)}\right) - \left(\frac{(\sim CE)}{(\sim C\sim E)+(\sim CE)}\right) \quad (17\text{-}1)$$

According to Equation 17-1, the different types of covariation information play a static role in assessing contingency. For example, all events of type CE play a role in increasing ΔP regardless of the context.

		Effect: Present	Effect: Absent
Cause	Present	CE	C~E
	Absent	~CE	~C~E

FIGURE 17-1 A traditional two-by-two contingency table used in models of causal reasoning.

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 281

It cannot be the case that some CE increases ΔP and some CE decreases ΔP.

Another causal induction model computes a measure called causal power from covariation information (Cheng, 1997; Novick & Cheng, 2004). For instance, the simple causal power for a generative cause is computed as follows:

$$\text{Power} = \frac{\Delta P}{\left(1 - \left(\frac{(\sim CE)}{(\sim C \sim E) + (\sim CE)}\right)\right)} \quad (17\text{-}2)$$

Again, it should not be difficult to see that, as in ΔP, the types of covariation information play static roles in computing an estimate of causal strength in the PowerPC theory.

Another dominant class of models is associative learning models, such as the Rescorla-Wagner (RW) model (Rescorla & Wagner, 1972). In this model, the associative strength V on the nth trial for each cue is updated according to the following calculation:

$$\Delta V_n = \alpha\beta(\lambda - \Sigma V_{n-1}) \quad (17\text{-}3)$$

In this equation, λ is 1 when the outcome is present and 0 when the outcome is absent. The parenthetical quantity is the amount of error on the nth trial; the difference between the outcome (□)\edq3\ and the summed associative strength of the present cues (ΣV_{n-1}). The saliency of the cue and the outcome are represented by the positive quantities α and the β parameters, respectively.

Like other models of causal learning, the RW model treats all observations of a given type uniformly (see Wasserman, Kao, Van Hamme, Katagiri, & Young, 1996, for details). For instance, when encountering a CE event, the change in association strength of C is as follows:

$$\Delta V_{\text{Cause}} = \alpha\beta(\lambda_{\text{Outcome}} - \Sigma V_{\text{Cause+Context}}) \quad (17\text{-}4)$$

Given normal values for the necessary parameters, this quantity will be positive and increase the perceived strength of the relationship between the cause and effect. Similarly, for all C~E events, the change in association is as follows:

$$\Delta V_{\text{Cause}} = \alpha\beta(\lambda_{\text{NoOutcome}} - \Sigma V_{\text{Cause+Context}}) \quad (17\text{-}5)$$

This quantity will be negative, leading to a decrease in the perceived strength of the relationship between cause and effect.[1]

Unlike the assumption shared by the discussed models, we believe that each type of evidence in Figure 17-1 is open to multiple causal interpretations. For instance, when C is present and E is absent, then it may be because there is a negative relationship between C and E, because there is no relationship between the two, or because C is indeed a cause of E but some necessary precondition was not satisfied. Similar interpretations can be made for each type of contingency information (see Table 17-1).

The decision about which interpretation to use is presumably influenced by multiple factors. One plausible influence is the reasoner's belief about the causal relationship at the time the observation is encountered. Of course, such beliefs should in turn depend on previously encountered observations. For example, if prior observations lead a reasoner to believe in a positive relationship between two variables, the reasoner may be more likely to interpret subsequent events according to the positive interpretation column in Table 17-1. If someone else believes that there is a negative relationship between events, that person may make interpretations more like those in the negative interpretation column.

In this chapter, we review three sets of studies, demonstrating that people spontaneously treat the same type of evidence differently because of beliefs developed during prior causal induction. In particular, our major thesis throughout these studies is that people develop hypotheses about causal relations early during causal learning and interpret subsequent data in light of these hypotheses. As a result, the reasoner's working hypothesis can then lead to identical data playing different roles. Such dynamic interpretations of data result in the primacy effect in causal learning, inferences about unobserved, alternative causes, and the context effect in interpretations of ambiguous stimuli.

Primacy Versus Recency Effects in Causal Induction

The use of an existing hypothesis in the interpretation of contingency information has great implications for the evaluation of sequential information. If a reasoner is given a set of evidence that suggests that C causes E, subsequent negative information may be reinterpreted in one of the ways shown in Table 17-1. For example, a piece of C ~E evidence may be interpreted as an

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 282

TABLE 17-1 Possible Interpretations for Each Type of Evidince

Evidence	*Positive Interpretation*	*No Relslion Interpretation*	*Negative Interpretation*
CE	C caused E, so E occurred when C occurred	C has nothing to do with E, and C and E just happened to occur together	C suppresses E, but something went wrong, so C and E occurred together
~C~E	C causes E, and E did not occur because C did not occur	C has nothing to do with E, and it just happened that E did not occur when C did not occur	C suppresses E, and E did not change because C did not occur
C~E	C causes E, but E did not occur because something went wrong	C has nothing to do with E	C spppresses E, so E did not occur because C occurred
~CE	C causes E, but E was caused by something else in this case	C has nothing to do with E	C suppresses E, so E occurred because C did not occur

example of the failure to satisfy a precondition, and a piece of ~CE evidence may be seen as the presence of an alternative cause. Conversely, on first encountering a majority of negative evidence at the beginning of an information set, people would interpret CE as a spurious correlation and ~C~E as an absence of any relationship.

Dennis and Ahn (2001) tested the prediction that the order in which people encounter evidence would influence causal strength estimates because people initially develop different hypotheses, which result in different interpretations of subsequent covariation information. Participants observed a sequence of trials, each describing the presence or absence of two events, and then judged the causal strength between the two events at the end of the sequence. Participants in the positive-first condition observed the bulk of the positive evidence, followed by the bulk of the negative evidence without any explicit marking for when the second set began. In the negative-first condition, participants observed the bulk of the negative evidence followed by the bulk of the positive evidence. Although the order was different, all participants observed ΔP of 0.

The PowerPC theory (Cheng, 1997) does not predict an order effect. Contingency-based models calculate the causal strength of an event over all available trials at once when enough observations are accumulated. Order of information does not change the probabilities used in Equation 17-2. Therefore, the ordering of information in Dennis and Ahn (2001) should have no effect on estimates of causal strength according to these models.

In the RW model, the strength of association between cue and outcome is updated at each trial, making the model sensitive to the sequence in which a series of learning trials is presented. In our simulation of the Dennis and Ahn (2001) experiments, the RW model yielded clear recency effects for almost every logical combination of parameters. The RW model predicts the recency effect because the degree to which an outcome is surprising determines associative learning. Evidence suggesting a positive relationship would be more surprising after a bulk of negative evidence (negative-first condition) than in the absence of such negative evidence (positive-first condition). Similarly, negative evidence is more surprising after presentation of positive evidence (positive-first condition) than in the absence of such evidence (negative-first condition). Consequently, in both conditions, the later information is more surprising and hence has a larger impact on associative strength, resulting in the recency effect.

Unlike the predictions of these two models, Dennis and Ahn (2001) found a primacy effect: Participants in the positive-first condition gave much higher estimates than those in the negative-first condition. Given that the primacy effect can pose a critical problem for all existing models of causal induction and that some research has found the opposite order effect of recency (Collins & Shanks, 2002; López, Shanks, Almaraz, & Fernández, 1998), it is crucial to understand the conditions under which the primacy effect occurs. Dennis (2004) and Marsh and Ahn (2005b) examined two possible reasons for obtaining the recency effect in causal induction.

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 283

One possible methodological difference between studies finding recency and those finding primacy is the frequency with which estimates of causal relations were made by participants. López et al. (1998) asked participants to estimate causal strengths multiple times during learning (step-by-step estimates), whereas Dennis and Ahn (2001) asked for the estimate only at the end of learning (end-of-sequence estimates). Hogarth and Einhorn's (1992) analysis of tasks used in the impression formation literature found that end-of-sequence estimates induced primacy, whereas step-by-step estimates tend to induce recency. Hogarth and Einhorn explain that impression formation involves belief updating, and the first piece of evidence (or an amalgamation of the first few pieces) serves as the anchor in end-of-sequence tasks. The anchor then serves as the light by which all other information is updated, resulting in a force toward primacy. On the other hand, in step-by-step tasks, people are forced to revise their hypotheses whenever they generate a new estimate, and thus the first piece of evidence no longer serves as an anchor after an estimate is made. According to Hogarth and Einhorn's belief-updating model, the weight of each new piece of evidence is adjusted based on a mechanism similar to the RW model: The more new information differs from the current position, the more weight it receives. Therefore, as in the RW model, the recency effect is predicted for step-by-step estimation.

Collins and Shanks (2002) presented a more direct investigation of the effects of estimate frequency on causal induction judgments. Using the same paradigm as in Dennis and Ahn (2001), they found that an end-of-sequence estimation procedure created a primacy finding, whereas more frequent estimation (every 10 trials) produced recency. Careful examination of Collins and Shanks's procedure produces reasons to suspect that the recency effect might have been induced by demand characteristics of the frequent estimate condition. Asking for an estimate only at the end of a sequence may implicitly cue participants to integrate over all information they have seen to make an estimate (e.g., "Because I have not been asked to make an estimate yet, then they must want me to use all of the information I have seen"). However, when some type of judgment is asked for every 10 trials, participants can interpret this as a cue that only the current information should be used (e.g., "Because I am repeatedly asked to make an estimate, there must have been some change in the data that warrants only using the most recent information"). Collins and Shanks failed to take cautionary measures to prevent this type of misinterpretation, as has been done in other studies (e.g., Catena, Maldonado, & Candido, 1998). Collins and Shanks compounded this problem by instructing participants that they will improve over the course of the experiment ("Although initially you will have to guess, by the end you will be an expert!" p. 1147).

Dennis (2004) replicated the work of Collins and Shanks (2002) with two critical modifications. Participants were asked to make causal strength estimates at every trial. This frequency was the strongest possible manipulation for the frequency of estimation argument, but at the same time, it reduces the demand characteristics in that there is no incremental set of data that participants might think as more important. Second, at each judgment, participants were asked to consider all data they had seen so far. Adding these instructions favors neither the recency nor the primacy effect, but eliminates the potential demand characteristics. With these two measures taken to eliminate the demand characteristics, Dennis found a strong primacy effect despite the fact that participants had to make step-by-step judgments.

A second possible explanation of recency effects is task complexity. López et al.'s (1998) learning materials were much more complex than those of Dennis and Ahn (2001). López et al.'s participants received information about a disease X and three possible symptoms. In one half of the learning sequence (the contingent block), one of these symptoms (A) was always paired with another (B). When the compound symptoms AB were presented, the patient usually had the disease, but when symptom B occurred alone, the disease was usually not present. This pairing suggests B was a worse predictor of the disease than was A. In the other half of the sequence (the noncontingent block), Symptom A was paired with a new cue (C). In this block, however, the disease occurred as often with C alone as with Compound AC, suggesting that Symptom C was a better predictor than A. Comparing the two blocks, higher ratings of the relationship between Symptom A and the disease should be given for the contingent than the noncontingent block. The order of these two blocks was manipulated to create different conditions in which either the contingent block was presented first or the noncontingent block was first. López et al. found that ratings

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 284

of the relationship between Symptom A and the disease were higher in conditions in which the contingent block was last (hence a recency effect).

López et al.'s (1998) design becomes complicated in that they simultaneously presented another set of stimuli with the same structure during the same learning phase (e.g., two contingent blocks each instantiated in its own disease). Furthermore, participants were simultaneously presented with two more sets of materials for the opposite order condition, thereby leading to symptom information for four diseases presented simultaneously. As a way to approximately illustrate complexity of the task given in this experiment (albeit in a somewhat arbitrary way), we can say that across all four sets of materials, participants had up to 20 hypotheses (3 single cues and 2 configural cues in each set times 4 sets) to keep track of by the time they got to the end of the experiment. Under this situation, the recency effect is more likely because participants would lose track of the hypotheses that they were testing and base their judgments on the most recent evidence. Participants could also fail to develop any hypotheses until later trials. In contrast, Dennis and Ahn's (2001) participants kept track of only one hypothesis. Marsh and Ahn (2005b) propose that the recency effect found in López et al. (1998) is an artifact of an overly complex procedure. Using the identical stimuli and procedure as López et al. but reducing the number of hypotheses to be tested to 5, we found a strong primacy effect (Marsh & Ahn, 2005b, Experiment 1).

In another experiment, Marsh and Ahn (2005b) doubled the cognitive load during learning by using two sets of stimulus materials, such that the number of causal relations to be considered would be 10. (This still would be half the amount López et al. found in 1998.) Given this increase, neither a recency nor a primacy effect was found. In this study, participants' spatial and verbal working memory capacity was also measured. We believe that the primacy effect did not occur in López et al.'s study because there were too many hypotheses to be examined early. If a subject has a large working memory capacity, then that subject would be more likely to be able to handle keeping track of so many hypotheses and be more likely to show the primacy effect. Indeed, Marsh and Ahn (2005b) found that participants' verbal working memory capacity positively correlated with the amount of primacy effect.

To summarize, work on order effects suggests that basic covariation information can be interpreted differently over the course of learning. As was indicated by results from Dennis and Ahn (2001), Dennis (2004), and Marsh and Ahn (2005b), information early in a learning sequence can greatly color the interpretation of later information. It is proposed that this early information serves as a basis by which hypotheses about causal relationships are formed. Later information is then differentially interpreted in overall estimates of causal strength. Such malleable interpretations are in contrast with the standard view of covariation data as uniform, static information.

Unobserved Alternative Causes

Models such as ΔP (Cheng & Novick, 1992), PowerPC (Cheng, 1997), and RW (Rescorla & Wagner, 1972) use information about whether an event occurred (such as in Figure 17-1) to evaluate the causal strength of variables. There are situations, however, in which such information is not available. Imagine a situation in which a doctor is treating a new disease. The doctor believes that there are two potential causes of the disease. The first is exposure to high levels of mercury, which can be measured using a blood test. The second is thought to be a genetic anomaly that is currently undetectable. In this case, one of the ostensible causes (the genetic anomaly) is unobservable. With respect to this cause, patients may be grouped as either having the disease or not, but they cannot be further broken down into the categories shown in Figure 17-1 (e.g., CE vs. ~CE and C~E vs. ~C~E). Such situations thus pose a problem for models that rely on contingency information as their input; these models cannot render judgments about the unobserved cause because the necessary input is not available.

Luhmann and Ahn (2003, 2005) demonstrated that, unlike the difficulties manifest in current models, people spontaneously make causal judgments about unobserved causes. The experimental situation used in these studies mirrors the example given in the preceding section. In each case, there were two causes and one effect. One of the causes was fully observable (similar to the mercury levels), and one was unobserved (similar to the genetic anomaly). The effect was always observable. In their Experiment 1, Luhmann and Ahn (2003) found that although they were allowed to withhold their judgments, participants were willing to make estimates on causal strength of unobserved causes. Because unobserved causes do not yield

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 285

covariation information, these findings imply that people used a different source of knowledge to make judgments about the unobserved causes.

A critical finding of Luhmann and Ahn (2003) was that observations of what the authors refer to as *unexplained effects* led to differential causal judgments of the unobserved cause. Unexplained effects are occasions on which the effect occurs in the absence of any observed cause. In the above example, unexplained effects would be an instance in which a person contracts the disease but tests negative for mercury. Luhmann and Ahn (2003) found that people judged the unobserved cause to be stronger when such instances of unexplained effects occurred than when they did not occur.

Unexplained effects imply the existence and operation of an unobserved alternative cause on those occasions, and that this belief triggers the creation of a hypothesis about the unobserved cause. Because unexplained effects indicate that the unobserved cause was responsible for the effect on that occasion, the hypothesis about the unobserved cause should initially posit a positive (i.e., generative) relationship of nonzero strength between the unobserved cause and the effect. This hypothesis is assumed subsequently to operate as described in the preceding section; that is, the hypothesis about the unobserved cause will color the interpretation of subsequent experience.

Specifically, our prediction was that if people maintain a hypothesis about the unobserved cause, then observations should be interpreted as confirming this belief (Lord, Ross, & Lepper, 1979). Thus, in conditions with unexplained effects, participants should believe the unobserved cause to be responsible for the target effect and thus interpret observations to conform to this belief. For instance, in the subsequent CE trials (i.e., joint presence of the observed cause and the effect), participants would believe that the unobserved cause is likely to be present as well, whereas in the subsequent ~C~E trials, they would believe that the unobserved cause is unlikely to be present. Next, we describe two experiments testing this prediction.

The first study sought to evaluate people's explicit beliefs about the presence or absence of the unobserved cause. To do so, we provided participants with a causal learning task like that used by Luhmann and Ahn (2003). To reiterate, one of the causes and the effect were fully observable, and the other cause was unobservable. In addition, after each trial, we explicitly asked participants to judge how likely it was that the unobserved cause was present on that occasion. Twenty-four participants viewed each of the four contingencies shown in Table 17-2 and made their likelihood judgments using an 11-point scale (0, definitely not pressed; 10, definitely pressed).

Table 17-3 shows mean ratings broken down by the four conditions and four trial types. The first finding to note is that the unobserved cause is most likely to be present during unexplained effects (~CE) trials, as expected. In the next analyses, we examined whether participants interpreted trials in light of their beliefs about the unobserved cause. If participants are interpreting observations as consistent with their beliefs about the unobserved cause, then they should believe that the unobserved cause covaries with the effect more in the two conditions with unexplained effects than in the two conditions without unexplained effects.

To test this, we compared CE trials and ~C ~E trials, the only trial types shared among the four conditions. If participants believed the unobserved cause covaried with the effect (which we predicted to be the case in the conditions containing ~CE), then participants should believe the unobserved cause to be more likely present on CE trials and more likely absent on ~C~E trials; the unobserved cause should covary with the effect. Participants who do not believe the unobserved cause covaried with the effect (which we predicted to be the case in the conditions without ~CE) may believe that the likelihood of the unobserved cause being present is more similar on these two trial types; the unobserved cause should not covary with the effect.

For each participant, their average rating for ~C ~E trials was subtracted from their average rating for CE trials. This composite score serves as an index of the degree to which participants believed the unobserved cause to vary with the effect. A 2 (C ~E present/absent)×2 (~CE present/absent) repeated measures analysis of variance was performed on this composite. This analysis revealed a significant main effect of unexplained effects (~CE trials), $F(1, 23)=8.77$, mean square error=84.52, $p<.01$, because the composite was higher in conditions that included ~CE (Mean [*M*]=3.60) than on conditions that did not include ~CE (Mean=1.72). This analysis suggests that unexplained effects not only lead to the perception of a stronger unobserved cause (as demonstrated by Luhmann & Ahn, 2003), but also led participants to

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 286

TABLE 17-2 Contingencies Used in Each of the Four Conditions

Condition	*~CE Present, C~E Absent*		*~CE Present, C~E Present*		*~CE Absent, C~E Absent*		*~CE Absent C, C~E Present*	
	E	*~E*	*E*	*~E*	*E*	*~E*	*E*	*~E*
Contigency structure	C 10	0	C 10	**10**	C 10	0	C 10	**10**
	~C **10**	10	~C **10**	10	~C 10	**10**	~C 0	10

Each condition contains CE and ~C~E observations. Only the presentation of ~CE and C~E observations differs.

believe that the unobserved cause covaries with the effect. Participants interpreted the unobserved cause differently on identical trials depending on whether they had observed unexplained effects.

Thus, we have demonstrated that unexplained effects result in the belief that the unobserved cause covaries with the effect and the belief that the relationship between the unobserved cause and the effect is strong. An obvious question is whether these beliefs are related. It seems plausible that people are able to make causal judgments about unobserved causes (and were confident in these judgments; see Luhmann & Ahn, 2003) because they have information about how the unobserved cause covaries with the effect. This information is not present in the input as current theories of causal learning assume but rather is imposed on the input by the reasoner. To explore whether people's trial-by-trial beliefs influence their subsequent causal judgments, a second experiment was conducted using a paradigm similar to that of Dennis and Ahn (2001) as described in the preceding section.

Fifty participants viewed each of two causal situations. The two situations used the set of trials represented in Figure 17-2. This set of trials was divided into two blocks, although there was nothing to indicate this to participants. One block contained unexplained effects, and the other did not. These two blocks could be ordered in one of two ways; the block containing unexplained effects could be presented either first or second. After viewing all trials, participants were asked to judge the causal strength of both the observed and unobserved cause. Note that, because the only manipulation was the order of the two blocks, participants always saw the same set of covariation data by the end of the sequence. Thus, any differences between conditions cannot be a result of the number or type of trials observed.

We predicted the order effect for the following reasons: When participants observed unexplained effects early, considerably more evidence was available to interpret as consistent with, and thus reinforce, the hypothesis about the unobserved cause. This additional reinforcement would lead participants to perceive the unobserved cause as a strong causal influence. When participants observed unexplained effects late in experience, a significant number of the observations had already occurred and thus were not interpreted in light of the newly formed hypothesis. This would lead participants to perceive the unobserved cause as a weaker influence.

As summarized in Figure 17-3, participants gave a significantly higher rating for the unobserved cause in the early unexplained effects condition (M=73.50, standard deviation [SD]=25.90) than in the late-unexplained-effects condition (M=61.66, SD=27.79),

TABLE 17-3 Average Trial-by-Trail Likelihood Ratings For the Various Trial Types in Each Condition Plus Marginal Averages

Condition	*~CE Present, C~E Absent*			*~CE Present, C~E Present*			*~CE Absent, C~E Absent*			*~CE Absent, C~E Present*		
	E		*~E*	*E*		*~E*	*E*		*~E*	*E*		*~E*
Likelihood (U)	C *5.8*		*5.80*	C *5.38*	**3.41**	*4.40*	C *4.1*		*4.10*	C *5.26*	**3.76**	*4.51*
	~C **7.51**	2.15	4.83	~C **7.84**	1.82	4.83	~C	3.16	3.16	~C	**2.7**	2.70

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 287

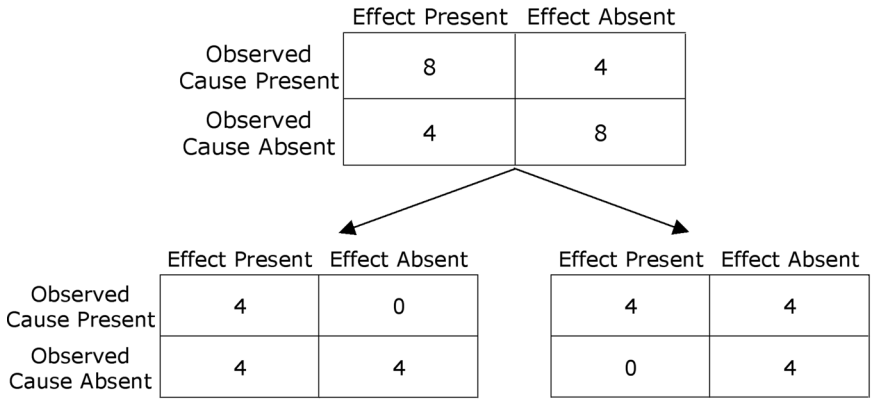

FIGURE 17-2 A summary of the observations presented to participants. The contingency table summarizes the covariation of the observed cause and the effect. The set of observations was divided into two blocks. One block contained unexplained effects; the other did not. The order of these blocks was manipulated.

$t(49)=2.89$, $p<.01$. These results support the idea that observations obtained after creating an unobserved cause hypothesis act to reinforce the hypothesis.

Note that this is not the only possible outcome. Theoretically, when participants in the late unexplained effects condition first observed an unexplained effect, they could create an unobserved cause hypothesis and reevaluate all previously obtained observations. Such retrospective reevaluation would likely require significant cognitive resources and thus may not be a generally economical strategy.

Consistent with the work of Dennis and Ahn (2001), these studies indicate that identical observations can be interpreted differently depending on the beliefs held by the observer. The two studies reported in this section demonstrate that such dynamic interpretations occur when evaluating unobserved causes just as they do with observed causes. Moreover, this differential interpretation influenced both explicit likelihood ratings and causal strength ratings. These findings suggest that unobserved causes are sometimes treated very much like observed causes. The observer can establish beliefs about an unobserved cause, interpret observations to overcome the absence of covariation information, and subsequently compute causal strength.

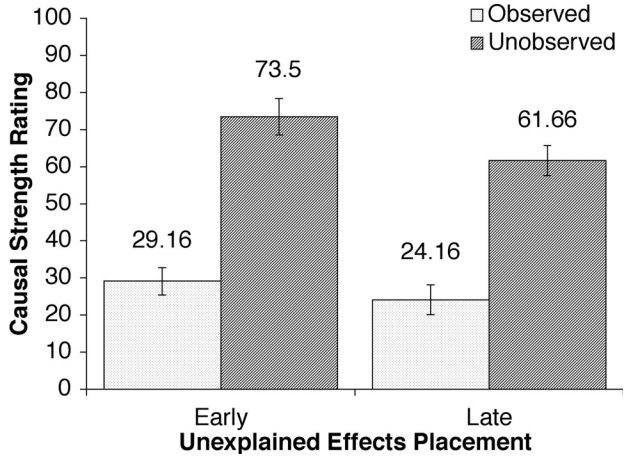

FIGURE 17-3 Participants' mean causal strength judgments. Error bars indicate standard error.

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 288

Context Effect in Interpretations of Ambiguous Stimuli

As discussed, the major models of causal reasoning (e.g., Cheng, 1997; Cheng & Novick, 1990; Rescorla & Wagner, 1972) deal with information that is uniformly presorted into the standard contingency table (see Figure 17-1). What occurs if information pertaining to a possible causal relationship is not so clearly defined regarding which of the four evidence types it represents? For example, consider trying to assess whether high stress causes insomnia. There are some events that are obviously instances of the presence of high stress (e.g., taking medical school entrance exams) and some events that are easily classified as the absence of high stress (e.g., sunbathing on a tropical island). However, there exists a wide spectrum of events between these two extremes that are not so clearly classified as the absence or presence of stress (e.g., waiting in a crowd, writing an e-mail, celebrating a milestone birthday). How would a reasoner assess a hypothesis such as "High stress causes insomnia" when the great amount of ambiguously identified evidence has no clear place in the classic representations of causal information?

Similar to how an existing hypothesis can affect the interpretation of sequential information and influence beliefs about an unobservable cause's operation, a governing hypothesis can be used to assess ambiguous causal information. For example, imagine that a reasoner keeps experiencing events in which high stress preceded a night of insomnia, whereas insomnia never followed a stress-free day. The reasoner then is asked to assess whether a friend was experiencing high stress given the observation that after a day spent reading the person did not sleep. The reasoner would use personal belief that high stress causes insomnia to interpret the observation as an instance of a high-stress day. Even though a day of reading in itself might not be stressful (e.g., reading the latest romance novel) or could be stressful (e.g., searching job postings in the classifieds), the given instance of reading would be classified as stressful because such an interpretation matches the reasoner's existing theory of the relation between events. In this way, information not inherent to the standard representation of covariation information would influence causal estimation.

To test the hypothesis that a governing causal hypothesis can cause an ambiguous event to be reinterpreted as a specific instance outlined in Figure 17-1, Marsh and Ahn (2005a) introduced ambiguous causal candidates into a traditional causal induction paradigm. Participants were presented with trial-by-trial evidence that indicated a strong covariation between two easily distinguishable, well-defined causal candidates and an outcome (see Figure 17-4 for actual trial frequencies). In one such sequence, for example, participants saw evidence that depicted bacteria that were of long (the candidate cause) as predominantly paired with the presence of nitrogen in soil samples (the effect event), whereas bacteria that were short were paired with the absence of nitrogen. To this basic paradigm, trials were added throughout the trial sequence depicting a candidate cause that was ambiguous regarding its membership in the cause-present or cause-absent class. In the previous example, these ambiguous trials would take the form of bacteria of intermediate length paired with the presence of nitrogen. The question of interest was whether participants would be willing to include this information in their assessments of causal information.

To address specifically if and how ambiguous information may be incorporated into evidence about causal relationships, Marsh and Ahn (2005a) in their Experiment 1 had participants report how many pieces of evidence they had observed by asking four questions that corresponded to the types of information found in Figure 17-1 (e.g., "On how many cases were the bacteria long and the nitrogen was present?" represented the CE cell). It was hypothesized that participants would use their current belief about the causal relationship between events to incorporate the ambiguous information into the traditional types of covariation evidence. For instance, if a participant believed that long bacteria (C) were generally associated with the presence of nitrogen in soil samples (E), then evidence depicting an ambiguous causal candidate (A) paired with the presence of nitrogen (i.e., a piece of AE evidence) would be interpreted as a piece of CE evidence. Under the same hypothesis, information that depicted the ambiguous cause in the absence of the effect (A~E) would be interpreted as evidence of type ~C~E. Therefore, ambiguous evidence of type AE should only be reflected in the CE estimate and likewise for ambiguous evidence of type A~E and the ~C~E estimate. Estimates of the two types of negative evidence should not be affected by ambiguous information because the negative evidence does not correspond to a way in which a hypothesis could be used to interpret ambiguous information. These predictions are depicted in Figure 17-4.

Participants in Marsh and Ahn's study (2005a) spontaneously assimilated ambiguous information into estimates of causal information. (See Figure 17-4 for mean

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 289

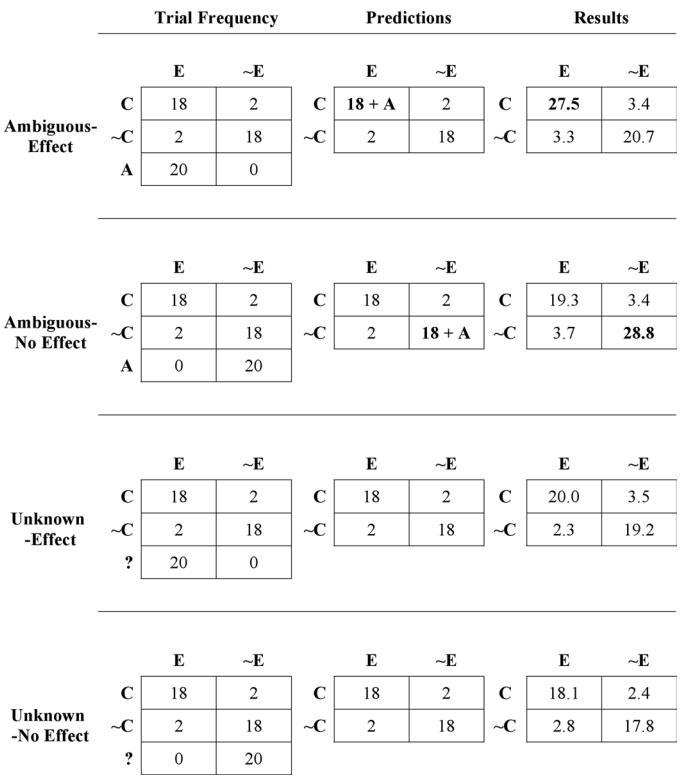

FIGURE 17-4 Number of trials for each type of evidence. C, cause; E, effect; A, ambiguous candidate cause; ?, unknown value of candidate cause.

estimates of the four types of evidence.) Specifically, if the ambiguous causal candidate was always paired with the presence of the target effect (ambiguous-effect), significantly more information was reported of type CE (M=27.5) than of type ~C~E (M=20.7), $t(19)=2.7$, $p<.02$. If the ambiguous causal candidate was always paired with the absence of the effect (ambiguous-no effect), more information was reported of type ~C~E(M=28.8) compared to type CE (M=19.3), $t(19)=3.3$, $p<.01$. In both of these conditions, there was no difference in the amount of information reported between the C~E and ~CE cells, $p>.20$. These findings are as predicted if participants were using an existing belief about the relationship between well-defined events to interpret ambiguous information.

The described results could have been found because participants felt forced by the constraints of the task to report ambiguous information instead of excluding this information or classifying it as a separate type of evidence from Figure 17-1. To guard against these possibilities, participants were not told ahead of the number or what frequency estimates they would make. In this way, participants could have represented ambiguous information as a separate class of information during the experiment with the expectation that they would be able to report it as such in final estimates. As a further precaution along these lines, the actual questions for estimation were presented one at a time in a random order. Ignorance of the complete spectrum of questions to be asked would have allowed participants to categorize ambiguous evidence as an additional category of information not represented in Figure 17-1. However, participants still incorporated ambiguous evidence.[2]

Another interesting finding is that the assimilation of ambiguous information was not complete, as can be noticed in Figure 17-4. If all of the trials illustrating the ambiguous candidate cause were assimilated by the governing causal hypothesis, then the key cells in Figure 17-4 should be closer to a total of 38 trials. It appears that only about half of the ambiguous trials that could have been incorporated into estimates were assimilated as such. These results seem

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 290

sensible: For ambiguous information to be assimilated by an existing hypothesis, said hypothesis must first be created. In the beginning of the experimental sequence, participants have yet to form a theory of how the possible causal candidates are associated with the effect. At that point, participants would not be able to informatively classify the ambiguous evidence. After observing information, the relationship between candidate cause and effect would have become clearer, allowing the formation of a causal hypothesis and subsequent interpretation of ambiguous evidence. Because pieces of evidence depicting ambiguous causes were sprinkled throughout the trial sequence, the early ambiguous evidence was experienced without the benefit of a governing hypothesis and therefore would not have been included in frequency estimates. As discussed with respect to unobserved causes, the cognitive load involved may be too great for the retrospective incorporation of ambiguous information. Therefore, only ambiguous information presented after the formation of a hypothesis could have been interpreted through the lens of the hypothesis and thereby included in frequency estimates.

\edq4\ In another study,\edq4\ we further examined if the strength of the hypothesized causal relation would moderate the amount of ambiguous information that is assimilated into causal judgments, such that the stronger the causal relation is believed to be, the more assimilation should occur. For example, if a person believes that high stress is always followed by insomnia, then any day with an ambiguous level of stress that was followed by insomnia should be counted as an example of high stress occurring with insomnia. However, belief in a weak relationship implies that every time high stress is present, insomnia does not necessarily follow. If a person believes that high stress is not necessarily followed by insomnia, then there is no principled reason to believe that every ambiguous stress day would occur with insomnia. This person should therefore be willing not to incorporate all ambiguous stress/insomnia days as examples of high stress and insomnia. In this way, the incorporation of ambiguous information would be mediated by the strength of the governing hypothesis so that less ambiguous information should be incorporated at weaker contingencies.

To test the effect of believed strength in causal relations, Experiment 2 of Marsh and Ahn (2005a) compared the treatment of ambiguous information in causal relationships of differing strengths. Four conditions were presented, each portraying a different strength relationship between the well-defined trials and the presence of the effect: a perfect relationship condition ($\Delta P=1.0$), a strong relationship condition ($\Delta P=0.6$), a weak relationship condition ($\Delta P=0.3$), and a no relationship condition ($\Delta P=0$). The exact trial frequencies used in this experiment can be seen in Figure 17-5. In all of the conditions, the ambiguous causal candidate always appeared with the presence of the effect, and there were 20 such ambiguous trials. We predicted that for the no relationship condition no preferential sorting of ambiguous evidence should occur because no hypothesis regarding the relationship between events could be formed. In contrast, for the three conditions for which a relationship existed between well-defined events (i.e., the perfect, strong, and weak conditions), the ambiguous evidence would be incorporated into the frequency estimate that matched the governing hypothesis, specifically the CE cell. Furthermore, the amount of assimilation would be a function of the strength of the relationship between the well-defined events.

Figure 17-5 depicts the results for the described experiment (Marsh & Ahn, 2005a). For the three conditions in which a relationship existed between the well-defined events, ambiguous evidence was incorporated into the CE frequency estimates as predicted. This finding is evidenced by significantly greater information reported in the CE cell than the ~C ~E cell for all three conditions, all t's>2.82, all p's$<.009$. Also as predicted, this difference was not significant in the no relationship condition, $p>.18$.

The results of this experiment also show that different amounts of assimilation were reported depending on the strength of the existing covariation relationship. By subtracting the ~C ~E estimate from the CE estimate for each condition, the amount of information that was preferentially sorted into the CE cell was calculated. These difference scores were then compared across conditions to see if more information was being sorted into the CE cell at different causal relational strengths. Figure 17-6 shows a graph of the mean difference scores for the four conditions. As the graph shows, the stronger the covariation between well-defined events, the more information was preferentially sorted into the CE estimate. Significantly more information was preferentially sorted in the CE estimate than the ~C ~E estimate in the perfect relationship condition ($M=7.8$) compared to the weak relationship condition ($M=3.4$). More information was also sorted preferentially in the perfect condition compared to the no relationship condition ($M=1.6$) and into the strong

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 291

Perfect Relationship

Trial Frequency	E	~E
C	20	0
~C	0	20
A	20	0

Predictions	E	~E
C	**20 + A**	0
~C	0	20

Results	E	~E
C	**28.3**	1.5
~C	3.6	20.9

Strong Relationship

Trial Frequency	E	~E
C	16	4
~C	4	16
A	20	0

Predictions	E	~E
C	**16 + A**	4
~C	4	**16**

Results	E	~E
C	**21.5**	5.5
~C	6.9	15.6

Weak Relationship

Trial Frequency	E	~E
C	13	7
~C	7	13
A	20	0

Predictions	E	~E
C	**13 + A**	7
~C	7	13

Results	E	~E
C	**15.3**	8.7
~C	8.3	12.4

No Relationship

Trial Frequency	E	~E
C	10	10
~C	10	10
A	20	0

Predictions	E	~E
C	10	10
~C	10	10

Results	E	~E
C	13.1	11.0
~C	11.4	11.7

FIGURE 17-5 Number of trials for each type of evidence for multiple relationship strengths. C, cause; E, effect; A, ambiguous candidate cause.

condition (M=6.7) compared to the no relationship condition, all t's>2.37, all p's<.03. In short, the strength of the believed causal relations greatly affects the amount of ambiguous information that is incorporated into estimates of causal information.

Having demonstrated the influence of the governing hypothesis on interpretations of ambiguous stimuli, an interesting question is whether people would assimilate *any* unknown information. That is, how robust is this phenomenon? To examine this issue, we created\edq5\ a condition in which the ambiguous causal candidate was replaced with an unknown candidate cause. More specifically, the unknown candidate was marked with a question mark and the word *unknown* appeared instead of a picture of the bacteria. In the instructions, participants were told that there was no information known about the candidate cause for these trials. Going back to our previous example of the relationship between high stress and insomnia, the unknown trials would be similar to a situation in which a reasoner has no information about the stress of the target person's day (or any information that can be used to infer the level of stress that day) and only learned that the person suffered from insomnia.

Note that, just like our ambiguous stimuli, an unknown candidate cause does not inform whether a given observation is positive or negative evidence toward a hypothesis. Thus, participants could also assimilate these trials in a way similar to how they assimilated ambiguous stimuli. For instance, if participants initially believed that the target effect is caused by a target cause, then participants could infer that an unknown cause paired with the presence of the effect must have been a case when the candidate cause was present, and when the unknown cause was paired with the absence of the effect, then the causal candidate must have been absent. This would be the most sensible inference to make if a person is forced to guess about the state of the unknown causal candidate. For instance, if high stress correlates with insomnia, on encountering a person who suffers from insomnia, it would be reasonable to guess that the person experiences a lot of high stress. Therefore, it is possible and even plausible in our experiments that people would spontaneously make inferences about unknown causal candidates as they did for ambiguous causal candidates.

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 292

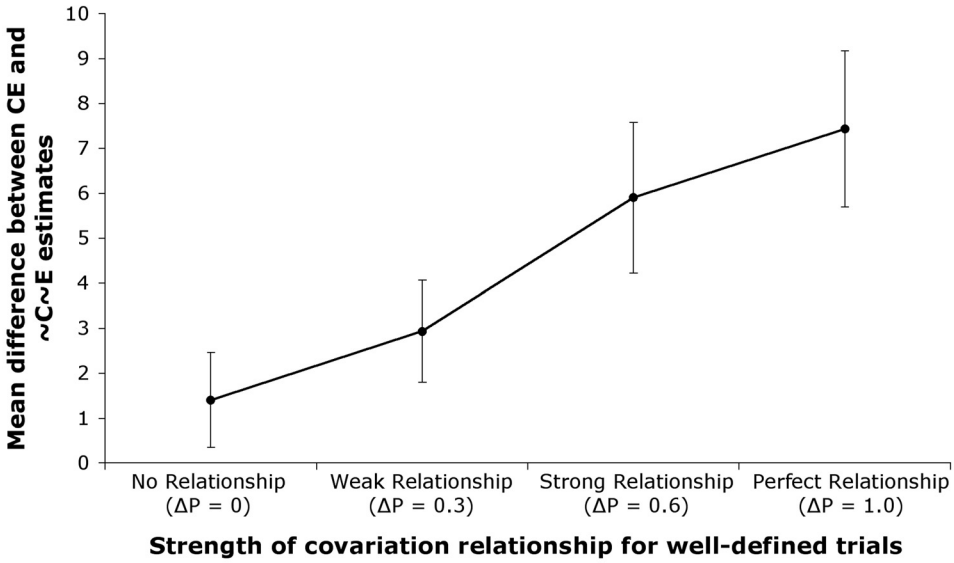

FIGURE 17-6 Mean difference between CE and ~C~E scores. Error bars indicate standard error.

In contrast, the mere presence of uncertainty might not be sufficient to provoke assimilation. Instead, ambiguous stimuli might need to be present for assimilation to spontaneously take place. The presence of ambiguous stimuli might trigger a need to classify the stimuli one way or another, which in turn results in assimilation. For instance, on encountering a bacterium with medium height, one might be enticed to determine whether it is long rather than simply leaving it as an undeterminable state and ignoring it.

To test the boundaries of which types of causal candidates would be incorporated into estimates of causal information, participants in Experiment 1 of Marsh and Ahn (2005a) were also given two additional conditions that contained unknown candidate causes. The representation of an unknown candidate cause was paired with the presence of the effect in one condition (unknown-effect) and the absence of the effect in the other (unknown-no effect). Because unknown information is lacking any type of structure on which a hypothesis can operate, it should not be incorporated into estimates of causal information. This prediction was validated in that estimates in the ambiguous conditions differed from the unknown conditions only in the cells predicted if ambiguous information was being incorporated via a governing hypothesis. That is, in the ambiguous-effect condition the CE cell mean estimate was greater than in the unknown-effect condition, and the ~C~E cell was greater in the ambiguous-no effect condition than the unknown-no effect condition. Furthermore, estimates for the four cells of the unknown conditions did not significantly differ from the number of well-defined trials presented (see Figure 17-4), demonstrating that participants can choose to exclude covariation information and are not bound to include it by the demands of the task.

We have demonstrated that people will incorporate information depicting an ambiguous causal candidate into their reports of relevant causal information. This was found despite the fact that participants had no advance knowledge of trial frequencies they would be asked to estimate. Participants could have excluded the ambiguous information in favor of waiting to classify it as a type of evidence not found in the classic representation of covariation information. Instead, participants spontaneously classified ambiguous causal candidates as the presence or absence of the candidate cause according to their governing hypothesis of the causal relationship.

Conclusion

As posited in the beginning of this chapter, we believe that covariation information can be dynamically interpreted during the process of causal learning. We have shown that such hypotheses can result in the discounting of later information in an information sequence,

Gopnik-CH_17.qxd 10/10/2006 8:08 PM Page 293

resulting in a primacy effect. We have also shown that causal hypotheses can alter the interpretation of covariation information into supporting an alternative, unobserved cause. In the last section, we likewise showed that an existing hypothesis could cause information that does not fit the normal representation of covariation information to be included into judgments of causal relationships. In particular, evidence that was ambiguous and normally has no place in the models of causal reasoning was reinterpreted by existing hypotheses and rendered usable.

The covariation-based models of causal induction have often been described as models of how people learn completely novel causal relations based on raw, untainted covariation data (see Tenenbaum and Griffith, this volume,\edq6\ for a more elaborated approach). We would argue that even when learning novel causal relations, people are driven to interpret covariation data in light of their own governing hypothesis. Such dynamic interpretations of covariation data are beyond the scope of existing covariation-based models.

References

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

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

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545–567.

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

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, *30*, 1138–1147.

Dennis, M. J. (2004). *Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships.* Unpublished doctoral dissertation, Yale University, Hartford, CT.

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

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3–19.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.

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

López, F. J., Shanks, D. R., Almaraz, J., & Fernández, P. (1998). Effects of trial order on contingency judgments: A comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 672–694.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality & Social Psychology*, *37*, 2098–2109.

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). *Learning about unobserved causes.* Manuscript submitted for publication.\edq7\

Marsh, J. K. & Ahn, W. (2003). Interpretation of ambiguous information in causal induction. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 775–780). Mahwah, NJ: Erlbaum.

Marsh, J. K. & Ahn, W. (2005a). *Ambiguity in causal learning.* Manuscript in preparation.\edq8\

Marsh, J. K., & Ahn, W. (2005b). *Order effects in contingency learning: The role of task complexity*. Manuscript submitted for publication.\edq9\

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

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

van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*, 127–151.

Wasserman, E. A., Kao, S.F., Van Hamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *Causal learning: The psychology of learning and motivation* (Vol. 34, pp. 208–264). San Diego, CA: Academic Press.

\edq1\Throughout, variables should be in italic. Should *C*, *E*, and *CE* be italic (as in equations)? Please indicate italics as needed throughout.
\edq2\Is italic *P* in ΔP correct?
\edq3\Is an open box correct for outcome? If not, provide.
\edq4\Provide reference for "another study."
\edq5\Indicate reference here (Marsh & Ahn, 2005a?).
\edq6\Cross reference to this volume. Indicate correct chapter.
\edq7\Update Luhmann & Ahn, 2005.
\edq8\Update Marsh & Ahn, 2005a.
\edq9\Update Marsh & Ahn, 2005b.