THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2007, 60 (3), 433–447

Psychology Press
Taylor & Francis Group

Predictions and causal estimations are not supported by the same associative structure

Miguel A. Vadillo and Helena Matute
Universidad de Deusto, Bilbao, Spain

Studies performed by different researchers have shown that judgements about cue–outcome relationships are systematically influenced by the type of question used to request those judgements. It is now recognized that judgements about the strength of the causal link between a cue and an outcome are mostly determined by the cue–outcome contingency, whereas predictions of the outcome are more influenced by the probability of the outcome given the cue. Although these results make clear that those different types of judgement are mediated by some knowledge of the normative differences between causal estimations and outcome predictions, they do not speak to the underlying processes of these effects. The experiment presented here reveals an interaction between the type of question and the order of trials that challenges standard models of causal and predictive learning that are framed exclusively in associative terms or exclusively in higher order reasoning terms. However, this evidence could be easily explained by assuming the combined intervention of both types of process.

Although the study of human contingency learning is far from being a young field in psychology (Jenkins & Ward, 1965; Smedslund, 1963; Ward & Jenkins, 1965) and was extraordinarily stimulated during the 1980s (Allan & Jenkins, 1983; Dickinson, Shanks, & Evenden, 1984; Wasserman, Chatlosh, & Neunaber, 1983), few would deny that this area of investigation is now experiencing one of its most intense moments. Many of the resulting studies have focused on the distinction between two general, theoretical views of causal learning. One of these frameworks assumes that causal learning is mostly determined by higher order cognitive processes related to statistical reasoning (Allan, 1980; Cheng, 1997; Cheng & Novick, 1992) or deductive inference (De Houwer, Beckers, & Glautier, 2002; Lovibond, Been, Mitchell, Bouton, & Frohardt, 2003). The second framework, on the contrary, regards causal learning as the result of rather mechanistic associative processes, which automatically capture interevent contingencies without the need for a deliberate and conscious process of reasoning (Allan, 1993; Dickinson et al., 1984). The distinction between these two frameworks (see Shanks, 2007, for a comprehensive review) is

Correspondence should be addressed to Miguel A. Vadillo or Helena Matute, Departamento de Psicología, Universidad de Deusto, Apartado 1, 48080 Bilbao, Spain. E-mail: mvadillo@fice.deusto.es or matute@fice.deusto.es

Support for this research was provided by Grant PI–2000–12 from Departamento de Educación, Universidades, e Investigación of the Basque Government to H.M. M.A.V. was supported by an FPI fellowship from the Basque Government (Ref. BFI01.31). We would like to thank Jan De Houwer, Francisco López, and Pedro Cobos for their insightful discussions concerning the experiments presented here.

http://www.psypress.com/qjep
DOI:10.1080/17470210601002520

to some extent isomorphic to the more general distinction between conscious rule interpretation and intuitive processing (Smolensky, 1988) and also to the distinctions between rational inferences and intuitive inferences (Hinton, 1990) or between associative processing and rule-based processing (Sloman, 1996).

One of the many strategies that have been followed to distinguish between the associative and the higher order cognitive reasoning accounts of causal learning consists in manipulating the type of question used to request participants' judgements about the perceived degree of relationship between a cue (or cause) and an outcome (or effect). From a strict associative point of view, this manipulation should have no impact on judgements. Regardless of the wording of the question, judgements should be based on the strength of the association between the cue and the outcome that has been learned during the sequence of trials in which both events have or have not co-occurred (see Cobos, Caño, López, Luque, & Almaraz, 2000). However, if higher order reasoning processes are responsible for participants' judgements, manipulating the test question may have an effect: Different questions might induce participants to think that different statistical indexes should be computed to solve the task efficiently. Empirical research has shown that the precise wording of the test question does have an influence on judgements (Crocker, 1982; De Houwer, Vandorpe, & Beckers, 2007; Gredebäck, Winman, & Juslin, 2000; Matute, Arcediano, & Miller, 1996; Matute, Vegas, & De Marez, 2002; Pineño, Denniston, Beckers, Matute, & Miller, 2005; Vadillo, Miller, & Matute, 2005; White, 2003).

For example, Matute et al. (1996) exposed their participants to a cue competition paradigm known as relative-validity (Wagner, Logan, Haberlandt, & Price, 1968; Wasserman, 1974, 1990) and studied the impact of using different test questions to assess the degree of cue competition. In the relative-validity design, participants in the experimental group are exposed to two types of trial. In one type of trial, two cues, A and X, appear in a compound followed by the presentation of the outcome (AX+). In the other type of trial, a different compound, but with a common cue, is followed by no outcome (BX−). For participants in the control group, both compounds are followed by the outcome in half of the trials and are followed by no outcome in the other half (AX+, AX−, BX+, BX−). Although Cue X is followed by the outcome with a probability of .50 in both groups, participants in the experimental group tend to judge the X−outcome relationship lower than do participants in the control condition. What Matute et al. (1996) found was that this relative-validity effect appears when subjects are asked to rate whether Cue X is a cause or an indicator of the outcome, but vanishes when subjects are asked to rate to what extent Cue X and the outcome co-occurred.

In a similar vein, Gredebäck et al. (2000) showed that other cue competition effects do also depend on the type of question. In their Experiment 1 they used a cue competition design in which during the first phase in some trials Cue A was followed by the outcome (A+), and in the remaining trials Cue C was followed by no outcome (C−). Then, during the second phase, participants saw two compounds of cues, each one containing one of the previously trained cues together with a novel cue, and both compounds were always followed by the outcome (AB+ and CD+). In this design, cue competition is observed if judgements for B are lower than judgements for D. Gredebäck et al. (2000) found a significant cue competition effect when participants were asked about the predictive value of the cue, as well as when they were asked about the causal relationship between the cue and the outcome. However, the cue competition effect did not reach statistical significance when participants were asked about the probability of the outcome given the cue, nor when they were asked about the frequency of cue−outcome pairings. Similarly, in Experiment 2 they showed that the conditioned inhibition effect disappears if the test question asks either for conditional probabilities or for the frequency with which the two events co-occur.

Other experiments have shown that the type of question also has an effect in designs different from those of cue competition. In a recent set of experiments, Vadillo et al. (2005) looked for differences between causal judgements (judgements about the strength of the causal relation between the cue and the outcome), predictive-value judgements (judgements about the predictiveness of the cue—that is, the value of the cue as a predictor of the outcome), and what they called prediction judgements (judgements in which participants have to estimate how likely it is that the outcome will occur).[1] On the one hand, they found that both causal and predictive-value judgements tended to be based on the cue–outcome contingency, as measured by the statistical index Δp—that is, the difference between the probability of the outcome given the presence of the cue and the probability of the outcome given the absence of the cue; that is, $p(o|c) - p(o|{\sim}c)$. On the other hand, they found that prediction judgements were based on the probability of the outcome given the cue—that is, $p(o|c)$.

As we have argued, the type-of-question effect is interesting from a theoretical point of view because it is beyond the scope of several models of causal learning. For instance, this effect cannot be explained by associative models, which assume that people rely on the strength of cue–outcome associations whenever they are asked to rate the degree of relationship between a cue and an outcome. However, although the question effect is problematic for a purely associative view of causal learning, it does not exclude the possibility that associative mechanisms play some role in the learning of the cue–outcome relationship. The only thing it shows is that there must be some nonassociative processes involved at least in the production of the response to different questions. In other words, associative processes might take part in what participants learn, even though they cannot wholly account for the flexible use that participants make of this acquired information (see Matute et al., 1996, 2002; Vadillo et al., 2005).

It is possible to develop an integrative view of causal, predictive-value, and prediction judgements that incorporates both associative mechanisms and reasoning processes. According to this view, some types of judgement can be regarded as being directly based on the strength of the cue–outcome association, while some others can be based on a combination of the cue–outcome and context–outcome associations. The decision of whether one or the other applies as a response to a given question is clearly nonassociative. However, the acquired knowledge that is used to construct that response would be, from this point of view, associative. For example, it is well known that the famous learning algorithm proposed by Rescorla and Wagner (1972) predicts that the asymptotic strength of the cue–outcome associations should be dependent on cue–outcome Δp (Chapman & Robbins, 1990; Danks, 2003; Wasserman, Elek, Chatlosh, & Baker, 1993).[2] Therefore, the judgements that have been shown to be dependent on Δp (causal judgements and predictive-value judgements; see Vadillo et al., 2005) could be a direct expression of the cue–outcome associative strength as computed by the Rescorla and Wagner (1972) learning algorithm. When making predictions, in contrast, participants should take into account all present cues that could be provoking the outcome, which means that both the cue–outcome and the

[1]In the prediction question used by Vadillo et al. (2005), participants had to estimate the likelihood of the outcome in the presence of the cue. Another possible prediction question would ask participants to predict the likelihood of the outcome in the absence of the cue (for an example of questions regarding what happens in the absence of the cue, see De Houwer et al., 2007). For the sake of simplicity, however, we always use the term prediction judgement as referring to what happens when the cue is present. Our assumption is that the results should, in general, be symmetrical for the cue-absent question.

[2] In fact, the predicted value of the asymptotic associative strength is exactly equal to Δp, if the learning rate parameter β is assumed to have the same value on outcome-present trials and on outcome-absent trials. When this constraint is met, the resulting algorithm is known as the restricted Rescorla–Wagner model (Lober & Shanks, 2000). If β has different values in outcome-present and outcome-absent trials, then the asymptotic value of the associative strength is no longer equal to Δp, but it is still dependent on this statistical index (higher levels of Δp lead to higher associative strengths).

context–outcome associations should be taken into account. It is easy to show that the asymptotic value of the addition of the cue's and context's associative strengths, as computed by the Rescorla–Wagner algorithm, is not dependent on Δp, but on $p(o|c)$, regardless of the saliences of the cue and the context (see Matute & Vadillo, 2005, for simulations illustrating this feature of the Rescorla–Wagner model). Thus, this would explain Vadillo et al.'s (2005) finding that prediction judgements, unlike causal and predictive-value judgements, are not dependent on cue–outcome contingency. Once it has been recognized that the effect of the type of question does not exclude the possible intervention of associative processes, the question is: Can we distinguish between this explanation, partly based on associative mechanism, and the alternative account exclusively framed in terms of higher order cognitive reasoning?

Although most higher order reasoning models do not usually include explicit algorithmic details (i.e., Cheng, 1997; Cheng & Novick, 1992; but see Cheng & Holyoak, 1995), it is commonly assumed that, from their point of view, what participants learn during the training experience is not associations between mental representations of the events, but some sort of mental model of a contingency table where the frequencies of each type of trial are stored. In situations where a single cue and a single training context are involved, it is assumed that participants store information about the frequency of four types of trial: trials in which both the cue and the outcome are present (type *a* trials); trials in which the cue is present but the outcome is absent (type *b* trials); trials in which the cue is absent but the outcome is present (type *c* trials); and trials in which both the cue and the outcome are absent (type *d* trials). An important feature of this view is that people are not supposed to keep information about the order in which they received these trials. That is, a participant having experienced 10 cue–outcome trials followed by 10 cue–no-outcome trials is supposed to acquire the same mental representation as that acquired by another participant having experienced 10 cue–no-outcome trials followed by 10 cue–outcome trials. Both of them should have stored a mental model with 10 Type *a* trials and 10 Type *b* trials.

Associative learning algorithms, on the contrary, are supposed to be highly sensitive to the precise order in which information was provided. Specifically, associative models generally assume that cue–outcome associations are constantly being updated as more information is provided. This means that the associative strength is strongly determined by the most recent contingencies. These models predict that, when contradictory information is received in different phases, what is learned in the last phase will overwrite what was learned previously, a process that has been called *catastrophic forgetting* (Hetherington & Seidenberg, 1989; Lewandowsky, 1991; McCloskey & Cohen, 1989; Ratcliff, 1990).

This differential prediction of higher order reasoning and associative accounts of causal learning is one of the best strategies that can be used to discriminate empirically between them (e.g., Chapman, 1991; López, Shanks, Almaraz, & Fernández, 1998). Thus, in the following experiment we use this trial-order strategy to assess the plausibility of a hybrid (partly associative and partly reasoning-based) account of causal and prediction judgements and test it against the explanation provided by standard higher order reasoning models.

Overview of the experiment

The major finding of Vadillo et al. (2005) was that prediction judgements were mostly determined by $p(o|c)$, whereas causal judgements were influenced by both $p(o|c)$ and $p(o|{\sim}c)$—that is, by Δp. Thus, predictions are only based on *a* and *b* trials, whereas causal judgements are based on *a*, *b*, *c*, and *d* trials. This means that manipulations affecting only *c* and *d* trials should have no impact on prediction judgements but should have an effect on causal judgements. Thus, if the order of *c* and *d* trials were manipulated while keeping the same order regarding *a* and *b* trials, prediction judgements should not be affected. But what pattern of causal judgements should we expect based on the different theoretical models? As we have just mentioned, from the point of view of higher order reasoning models, the order of trials should have no effect, and, therefore, causal judgements should be the same regardless of the order in which information about *c* and *d* trials were presented. On the contrary, associative models predict that varying the order of *c* and *d* trials should have an impact on causal judgements. Thus, according to our hypothesis, if causal judgements are based on the strength of the cue–outcome association as computed by the Rescorla–Wagner algorithm, then these judgements should be larger if most Type *d* trials are presented at the last phase of training than if most Type *c* trials are presented at that moment. Additionally, prediction judgements should not change depending on the order of *c* and *d* trials. This manipulation would affect the associative strength of the cue, but it would also affect the associative strength of the context in the opposite direction. Thus, if prediction judgements are based on the addition of the associative strength of the cue and the associative strength of the context, then manipulating the order of *c* and *d* trials should not have an effect on prediction judgements. It is only for causal judgements that we expect this manipulation to have some relevance.

Summarizing, both the associative account and the higher order reasoning account predict that the order of *c* and *d* trials should have no impact on prediction judgements. However, they make different predictions regarding whether such manipulation should have an effect on causal judgements. The associative account predicts a recency effect in causal judgements (i.e., causal judgements should be larger if most *c* trials are presented first, and most *d* trials are presented in the last training phase), whereas the higher order reasoning account predicts no trial-order effect.

These predictions were tested in this experiment with the design shown in Table 1. Two groups of participants were exposed to two phases of training. The probability of the outcome given the cue was set to .50 in both phases for both groups (15 *a* and 15 *b* trials were presented in each phase in both conditions). However, the probability of the outcome in the absence of the cue was different in each training phase. For participants in group CD, most *c* trials were presented during the first training phase, and most *d* trials were presented during the second training phase. This resulted in a higher probability of the outcome in the absence of the cue during the first phase, $p(o|{\sim}c) = .83$, obtained with 25 *c* and 5 *d* trials, than during the second phase, $p(o|{\sim}c) = .17$, obtained with 5 *c* and 25 *d* trials. For participants in group DC, the order of these blocks of trials was reversed: Most *d* trials were presented during the first training phase, $p(o|{\sim}c) = .17$, obtained with 5 *c* and 25 *d* trials, and most *c* trials were presented during the second phase, $p(o|{\sim}c) = .83$, obtained with 25 *c* and 5 *d* trials. Despite this critical difference in the order of trials, participants in both groups were exposed to the same overall $p(o|c)$ and $p(o|{\sim}c)$ (both equal to .50) and also to the same overall Δp (equal to 0). According to higher order reasoning models of causal induction this should give rise to equal causal judgements and predictions in both groups. However, according to our associative perspective, prediction judgements should be identical but causal judgements should be more sensitive to the cue–outcome contingency of the second training phase. Given that *c* trials should reduce the strength of the cue–outcome association and that *d* trials should increase it, this model predicts a greater judgement when most *d* trials are presented in the last training phase (CD condition) than when most *c* trials are presented in that phase (DC condition).

Method

Participants and apparatus

The experiment was run simultaneously over the Internet and in traditional laboratory conditions (for recent reviews on Internet-based research, see Birnbaum, 2000; Gosling, Vazire, Srivastava, & John, 2004; Kraut et al., 2004). Concerning the Internet replication, 59 anonymous Internet users volunteered to take part in this experiment. These participants were randomly assigned to

Table 1. *Design summary of the experiment*

	Phase 1				Phase 2			
Group	*Trial frequencies*	p(o|c)	p(o|∼c)	Δp	*Trial frequencies*	p(o|c)	p(o|∼c)	Δp
CD	15a, 15b, 25c, 5d	.50	.83	−.33	15a, 15b, 5c, 25d	.50	.17	.33
DC	15a, 15b, 5c, 25d	.50	.17	.33	15a, 15b, 25c, 5d	.50	.83	−.33

one of the two experimental conditions, which resulted in 30 participants in group CD and 29 participants in group DC. In order to comply with ethical regulations for human research over the Internet (Frankel & Siang, 1999) we decided not to record any data without the consent of the participants. Thus, the conditions in which these participants performed the experiment are completely unknown to us. However, we controlled for the potential noisy data obtained through the Internet by running an additional pool of 67 participants in our laboratory. These were psychology students from Deusto University who took part in the experiment voluntarily. Random assignment of these laboratory participants to the two experimental conditions resulted in 35 participants in group CD and 32 in group DC. These students performed the experiment in a large computer room, where adjacent participants were seated at about 1.5 m apart from each other and were exposed to different experimental conditions. This double-location procedure allows us to check that the effects under study can be generalized to less controlled conditions than those of the laboratory, while making sure that the potential noise introduced by the Internet-based methodology is not affecting the results.

The experimental program was an adaptation of the allergy task that we commonly use in experiments on human contingency learning (Bárcena, Vadillo, & Matute, 2003). This program was implemented in an HTML document dynamically modified with JavaScript, which allowed us to perform the experiment in any computer connected to the World Wide Web with a standard Internet browser.

Design and procedure

At the beginning of the experiment, participants were shown the following instructions on the computer's screen:

> Imagine that you are a specialist who wants to study to what degree the consumption of a medicine causes, as a secondary effect, an allergic reaction. The medical records of a series of patients will be presented. You will first see a card that tells you whether a patient has taken the medicine. Once you have read it, you will see, on a second card, whether the patient did or did not develop the allergic reaction. After that, you will see the cards for the next patient, and so on. After seeing all the patients' records, you will have to assess the relationship between the medicine and the allergic reaction.

After having read these instructions, participants were exposed to a sequence of 120 trials. Each trial began with the presentation of a medical card where it could be read whether or not that trial's patient had taken a medicine called Dugetil. Below this card and on the same screen, participants were asked to predict, giving a *yes/no* response, whether this patient would develop an allergic reaction to the medicine.[3] After having entered their response for that trial, participants could see in another medical card, which was presented underneath the yes/no

[3] Although it is well known that a high frequency of judgements can induce a recency effect (Catena, Maldonado, & Cándido, 1998), previous studies have shown that frequent yes/no responses do not affect participant's ratings given at the end of the experiment in a numerical scale (see Matute et al., 2002). For instance, both Collins and Shanks (2002) and Matute et al. (2002) observed an absence of recency in a numerical judgement given at the end of the experiment, in spite of participants being requested to give this yes/no responses during training. The recency effect only appeared when the numerical judgement itself was requested with a high frequency during training. Thus, it is unlikely that our yes/no responses are affecting ratings given at the end of the experiment.

buttons, whether or not the patient had actually developed the allergic reaction (see Bárcena et al., 2003). In each trial, the patient could take the medicine and develop the allergy (Type *a* trial), take the medicine but not develop the allergy (Type *b* trial), not take the medicine but develop the allergy (Type *c* trial), or not take the medicine and not develop the allergy (Type *d* trial). During the experiment, participants in both conditions were exposed to 30 trials of each type, but the distribution of these trials was different for each group: Participants in group CD were first exposed to a set of 15 *a*, 15 *b*, 25 *c*, and 5 *d* trials and then, without any visible interruption, to a set of 15 *a*, 15 *b*, 5 *c*, and 25 *d* trials; participants in group DC were exposed to these two sets of trials in the reverse order.

After having seen the whole sequence of 120 trials, a different screen appeared on which participants were asked two questions that appeared simultaneously, one above the other (with the position of question counterbalanced across participants). Participants could answer these questions in any order and were allowed to change their ratings before finishing the experiment. One of the questions can be translated as *To what extent do you believe that Dugetil is the cause of the allergic reaction?* Participants were asked to respond to this question by clicking any point in an scale numbered from 0 to 100, with the opposite ends labelled as *It is definitely not the cause* and *It is definitely the cause*. The second question can be translated as *If a patient has taken Dugetil, to what extent do you believe that this patient will develop the allergic reaction?* Participants were asked to respond to this question in a separate scale, also numbered from 0 to 100, but with the opposite ends now labelled as *Definitely will not develop it* and *Definitely will develop it*. After having entered their judgements, participants were allowed either to change their ratings or to finish the experiment.

Although causal judgements should theoretically be requested with a bipolar rating scale ranging from −100 to 100 (because normative measures of causality, such as Δp, can adopt negative values), we decided to use a unidirectional scale from 0 to 100 in order to keep consistency with prediction judgements, which can never adopt negative values (i.e., an event cannot occur with a negative probability). In addition, it should be noted that in the present experiment we were more interested in detecting which type of judgement (causal or prediction) was affected by the order of *c* and *d* trials than in studying the precise values of the judgements given by participants in response to each question. Moreover, the use of a unidirectional scale keeps the procedure as consistent as possible with our previous studies (Vadillo et al., 2005).

Results

As the following analyses confirm, data gathered in the laboratory and over the Internet were not different from each other. Thus, data from both samples were collapsed. Figure 1 shows mean prediction and causal judgements at test collapsing both sets of data. The order of trials seems to have had no effect on prediction judgements, but causal judgements, on the contrary, have been affected systematically by this variable: As expected, causal judgements are higher in group CD than in group DC, suggesting a recency effect for these judgements.

These general impressions were confirmed by a 2 (type of question: causal vs. prediction) × 2

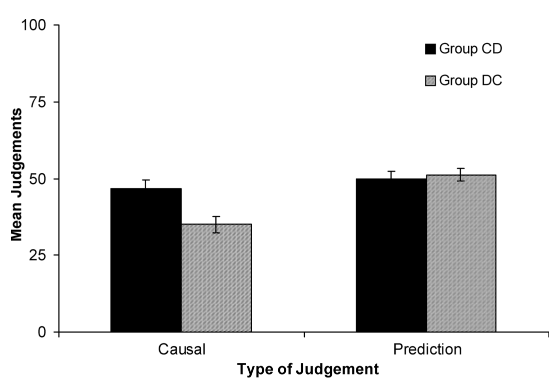

Figure 1. *Mean judgements at test given in response to the causal and prediction questions. Error bars represent the standard errors of the means.*

(group: CD vs. DC) × 2 (location: laboratory vs. Internet) mixed analysis of variance (ANOVA) performed on participants' judgements. This analysis yielded a main effect of the type of question, $F(1, 122) = 21.07$, $MSE = 282.49$, $p < .001$, and a marginally significant main effect of group, $F(1, 122) = 3.78$, $MSE = 465.95$, $p = .054$, as well as a significant interaction between type of question and group, $F(1, 122) = 9.6$, $MSE = 282.49$, $p < .005$. All main effects or interactions involving the location in which the experiment was run were nonsignificant, all Fs < 1; therefore, this factor was not used in subsequent analyses.

The interaction of the type of question and the trial order supports our hypothesis: Trial order effects are different for each question. A priori t test for independent samples confirmed that causal judgements were significantly higher in group CD than in group DC, $t(124) = 3.2$, $p < .01$, and that there was no significant difference between groups in prediction judgements, $t(124) = 0.39$, $p = .69$.

Discussion

These results indicate that the manipulation of the order of *c* and *d* trials gives rise to recency effects in causal judgements, whereas this manipulation has no effect on prediction judgements. In their present state, none of the most widely cited models of causal and predictive learning can explain this pattern of results without additional assumptions. On the one hand, the higher order reasoning account (Cheng, 1997; Cheng & Holyoak, 1995; Cheng & Novick, 1992) provides a partially satisfactory explanation for some of the evidence. For example, the lack of trial-order effects in prediction judgements is perfectly coherent with the predictions that one could draw from this theoretical perspective. Additionally, the fact that prediction and causal judgements follow different statistical indexes suggests that participants are aware of the differences between making predictions and assessing the strength of causal relations. This sensitivity to the subtle differences between the demands posed by a prediction question and those posed by a causal question is certainly in line with the hypothesis that people act as intuitive statisticians and possess some abstract knowledge about statistical relations (Buehner & May, 2003; Waldmann, 2000; Waldmann & Holyoak, 1992). However, given that the order in which information is presented is not considered relevant (or given, at least, that its relevance has not been addressed) in these models, they remain unable to explain the trial-order effects that we observed in causal judgements (see also López et al., 1998).

On the other hand, it is equally difficult to propose a complete explanation of our results from a purely associative perspective. Associative mechanisms (e.g., Dickinson & Burke, 1996; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994) could easily account for the recency effects observed in causal judgements, but they cannot explain the absence of recency in prediction judgements that was observed in the present study. Moreover, if judgements about a cue–outcome relationship are always based on the strength of the cue–outcome association then causal and prediction judgements should be equivalent, and manipulations affecting one of them should also affect the other (Cobos et al., 2000). This prediction is not supported by our data.

It seems, therefore, that a successful explanation of the present results cannot be framed exclusively in terms of higher order cognitive reasoning processes or in terms of associative processes. However, one need not conclude from this that both theoretical perspectives should be abandoned. Instead of that, one could simply regard them as incomplete but compatible accounts and try to find an integrative explanation borrowing concepts and ideas from both sets of theories. The hybrid framework that provided the starting point for this experiment is one of such possible accounts. According to our hypothesis, participants' basic learning of the cue–outcome relationship could be associative, but the subsequent use that participants make of this associative knowledge can be determined by higher order cognitive processes similar to those invoked by statistical

(Cheng, 1997; Cheng & Holyoak, 1995) and inferential (De Houwer, 2002; De Houwer & Beckers, 2002; De Houwer et al., 2002; Lovibond et al., 2003) reasoning accounts of human causal learning. In other words, participants can learn associations and use these associations in a more flexible way than that assumed by associative theories.

As we have discussed in the Introduction, one way of materializing this general perspective is by assuming that participants learn the cue–outcome contingency through an associative learning mechanism similar to the one proposed by Rescorla and Wagner (1972). Then, causal judgements are directly based on the strength of the cue–outcome association, whereas prediction judgements are based on the sum of the associative strengths of all the cues (including the context) that could cause the outcome at test. When $p(\text{o}|{\sim}\text{c}) > 0$, participants will probably perceive the context as a relevant causal factor, which can potentially contribute to the occurrence of the outcome. Thus, an accurate prediction of the outcome at test should be based not only on what is known about the cue (i.e., the cue–outcome association), but also on what is know about the context (i.e., the context–outcome association). In the Introduction, we have argued that, in situations involving the four trial types (*a*, *b*, *c*, and *d*) here manipulated, the summed associative strengths of the cue and the context are at asymptote equivalent to $p(\text{o}|\text{c})$ and provide, therefore, an appropriate basis on which participants could base their prediction judgements.

This formal account of causal and prediction judgements provides a straightforward explanation of our pattern of results. A simulation of the conditions used in our groups CD and DC is shown in Figure 2. As can be seen there, if causal judgements are assumed to be based on the associative strength of the cue, as computed by the Rescorla–Wagner learning algorithm, then a strong recency effect is expected for these judgements; that is, causal judgements should be higher in condition CD than in condition DC. However, prediction judgements are assumed to be based on the sum of the associative strengths

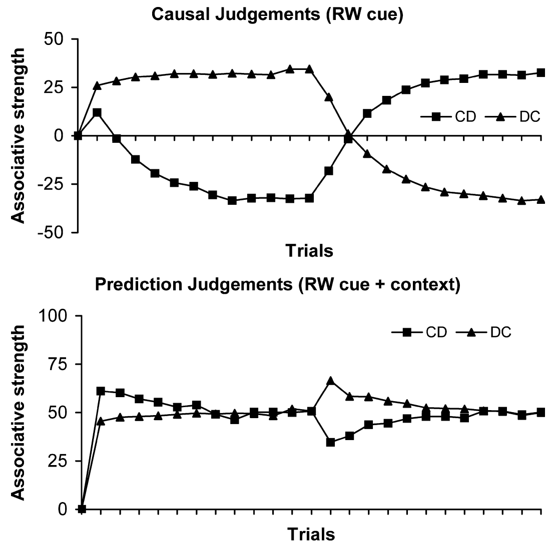

Figure 2. *Simulation of the Rescorla–Wagner model showing both the predicted associative strengths of the cue and the sum of the cue's and the context's associative strengths for conditions CD and DC. Learning rate parameters were assigned the following values:* $\alpha_{Cue} = .8$, $\alpha_{Context} = .5$, $\beta_{Outcome} = .6$, *and* $\beta_{NoOutcome} = .6$. *For each condition, 10,000 iterations with randomized trial orders within each phase were performed.*

of the cue and the context, and this sum is predicted to be similar for both groups, with only minor preasymptotic divergences in both phases. Thus, were prediction judgements based on these summed associative strengths, judgements would be equal in both trial-order conditions.

Note that the articulation of associative and reasoning processes could also be achieved by alternative formalizations. Here we have discussed how the present results could be explained assuming that the process of learning depends on an associative algorithm similar to the one described by Rescorla and Wagner (1972). However, we could also explain our results by assuming that learning is based on a simpler and more parsimonious associative mechanism like the one proposed by Bush and Mosteller (1951) or Hebb (1949). For example, the learning algorithm proposed by Bush and Mosteller (1951) is sensitive to $p(\text{o}|\text{c})$ but not to Δp. If people learnt the cue–outcome association using such mechanism, then they would be able to make accurate predictions (though not accurate causal judgements) basing their judgements directly on the strength of the cue–outcome association. However, participants could easily give an accurate assessment of the causal relation by subtracting the associative strength of the context–outcome association from the associative strength of the cue–outcome association. This formalization, framed in terms of the Bush–Mosteller learning algorithm, is similar to some current accounts of flexibility in human causal judgements based on the comparator hypothesis (Miller & Matzel, 1988; see, for example, Beckers, De Houwer, Pineño, & Miller, 2005; Pineño et al., 2005) and is also similar to an earlier associative account of human causal learning proposed by Shanks and Dickinson (1987), according to which people learn an association between the cue–context configuration and the outcome and also an association between the context-alone configuration and the outcome; causal judgements would be based on the difference between these two associations.

Our explanation based on the Bush–Mosteller learning algorithm bears some resemblance to the one framed in terms of the Rescorla–Wagner learning model. In the account based on the Rescorla–Wagner model, causal judgements are a direct expression of the cue–outcome association, and prediction judgements are the sum of the cue–outcome and the context–outcome associations. However, in the account based on the Bush and Mosteller learning algorithm, prediction judgements are a direct expression of the cue–outcome association, and causal judgements subtract the strength of the context–outcome association from the strength of the cue–outcome association. With our present data we cannot favour one of these explanations over the other.

In the present paper we have focused on the effect of the type of question in contingency designs where only four types of trial are used (resulting from combining the presence or absence of the cue with the presence or absence of the outcome). However, researchers have observed that the type of question also affects cue competition effects. Usually, cue competition is stronger if participants are asked to assess the causal relationship between the cue and the outcome than if they are asked to predict the outcome based on the presence of the cue (Gredebäck et al., 2000), if they are asked to rate the degree with which the cue and the outcome co-occur (Matute et al., 1996), or if they are asked to rate the cue–outcome predictiveness (Gredebäck et al., 2000; Pineño et al., 2005). Cue competition is also stronger if cues and outcomes are presented as causes and effects than if they are described as predictors and outcomes (De Houwer et al., 2002; Pineño et al., 2005). These results are not easy to explain from our hypothesis based on the Rescorla–Wagner algorithm. We have proposed that causal and predictive-value judgements are based on the associative strength of the target cue, whereas predictions are based on the combined associative strengths of the target cue and the context. However, the joint associative strength of a blocked cue and the context will always be lower than the associative strength of a nonblocked control cue and the test context. In other words, blocking should be observed regardless of whether participants base their judgements on the associative strength of the target cues or on the combination of the associative strengths of the target cues with the context. Thus, blocking should be observed even if the test question (e.g., prediction question) or the instructions (e.g., predictive scenario) induce participants to combine the associative strengths of the cue and the context. In spite of this, the addition of the associative strength of the context might in some situations reduce the relative difference between the response to the blocked cue and the response to its corresponding control cue: If a blocked cue receives less associative strength than the control cue, then the blocked cue and the context together will also have less associative strength than the control cue and the context together, but their relative difference will be smaller. Thus, a reduction of blocking is expected if participants combine the associative strengths of the target cues with the associative strength of the context.

Although we have argued that the recency effect observed in causal judgements is difficult to explain by current higher order reasoning models, it could perhaps be explained from their perspective by assuming that participants store not only a mental representation of a contingency table during training, but also more detailed information about the order in which this information was provided. This alternative coding strategy would allow participants to selectively focus on the most recent information when making their causal ratings. Although this explanation may be valid for some simple experimental designs (e.g., acquisition–extinction), it is not clear whether participants can store information in that format in complex designs like the one used in this experiment, in which there are all types of trial in both training phases, and the trial-order manipulation refers only to *c* and *d* trials (which are probably less salient than *a* and *b* trials). And if participants are able to do so, it remains unclear which can be the nature of the learning process responsible for the formation of such mental representation. Additionally, in the absence of a formalized model incorporating these hypotheses one cannot easily make a priori predictions of how participants' behaviour would be in situations like those used in our procedure or in alternative paradigms. In fact, the development of such formalization is one of the most important future challenges of higher order reasoning models of causal induction (De Houwer, Beckers, & Vandorpe, 2005). Thus, although we cannot exclude the possibility that our results would be one day entirely explainable in terms of higher order processes alone, we prefer to emphasize their consistency with associative models until a more scientifically testable account of trial-order effects is offered in terms of reasoning.

In the present paper we have focused on the distinction between causal and prediction judgements. Of course, our analysis could be extended to other types of judgement. For example, Vadillo et al. (2005) made a distinction between prediction judgements (asking participants to predict the likelihood that the outcome will occur) and predictive-value judgements (asking participants to assess whether the cue is a good predictor of the outcome) and found that, unlike prediction judgements, predictive-value judgements tended to be based on Δp. They concluded that the explanation given for causal judgements could be equally valid for predictive-value judgements. De Houwer et al. (2007) have also explored an alternative type of judgement, which they called preparatory judgements, in which participants were asked to prepare for the occurrence of the outcome if they thought that it would occur in a given trial. De Houwer et al. showed that preparatory judgements when the cue was presented (i.e., cue-present preparatory judgements) were less sensitive to Δp than were causal judgements, whereas they were strongly influenced by $p(o|c)$. Similarly, preparatory judgements in the absence of the cue were mostly determined by $p(o|{\sim}c)$. The similarities between the results observed when prediction questions and preparatory questions are used could be indicating that the mechanisms underlying prediction judgements are probably at the basis of preparatory behaviour as well.

Some of the results observed in the present experiment seem to contradict the results of Matute et al. (2002), who showed (a) that prediction questions are more sensitive to recency effects than are causal questions and (b) that requesting participant's judgements only at the end of the experiment produces an integration of all the information received or, in other words, prevents recency effects. Our results contradict (a) because in this experiment we observed a recency effect that affected only causal judgements (no recency was observed in prediction judgements), and they also contradict (b) because we observed this recency effect in causal judgements even though participants were only requested to give a single judgement at the end of the experiment. The absence of recency in prediction judgements in the present experiment is not really problematic: If these judgements are based solely on *a* and *b* trials, then there is no reason to expect recency in prediction judgements, unless the order of *a* and *b* trials is manipulated. Consistent with this, we did not observe trial-order effects in prediction judgements when only the order of *c* and *d* trials was manipulated. Therefore, with respect to (a), it is not surprising that causal judgements were more sensitive to recency effects than were prediction judgements in the present experiment.

An important difference between the present experiment and those reported by Matute et al. (2002), which could potentially account for the divergent results with respect to (b), is the degree of difficulty of the task used in both series of experiments. Whereas Matute et al. (2002) used a very simple acquisition–extinction design (see also Collins & Shanks, 2002, for a similar result using a slightly more complex trial-order manipulation), in the present experiment we used a rather complex design where the transition from the first to the second phase was relatively difficult to perceive and remember. Several researchers have noted that simpler processing occurs as the difficulty of the task increases (e.g., see Le Pelley, Oakeshott, & McLaren, 2005). Thus, quite possibly, the complexity of the task used in this experiment induced participants to enrol in simpler processing than that in Matute et al.'s (2002) experiments. An additional difference between both sets of experiments is that Matute et al. (2002) requested prediction judgements by presenting the question during a training or test trial (the cue and the target question were presented simultaneously), whereas we requested judgements by means of a questionnaire that was presented at the end of the experiment. This procedural difference might have also contributed to our finding recency with a single judgement at the end of the experiment.

We would not like to end the present paper without remarking on the similarities between the results observed in the laboratory and those observed over the Internet. The lack of a main effect of location or an interaction involving this factor indicates that the results were essentially the same for the Internet sample and for the laboratory sample, a result that confirms previous findings of other studies that also show a clear correspondence between traditional methodologies and Internet-based research (e.g., Buchanan & Smith, 1999; McGraw, Tew, & Williams, 2000; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Vadillo et al., 2005). These consistent similarities between experiments performed in the laboratory and experiments performed over the Internet encourage the still infrequent use of online experiments in our area.

REFERENCES

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgement tasks. *Bulletin of the Psychonomic Society, 15*, 147–149.

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin, 114*, 435–448.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation, 14*, 381–405.

Bárcena, R., Vadillo, M. A., & Matute, H. (2003). *Allergy task for Windows* [Computer program]. Available from http://www.labpsico.com

Beckers, T., De Houwer, J., Pineño, O., & Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 238–249.

Birnbaum, M. H. (Ed.). (2000). *Psychological experiments on the Internet.* San Diego, CA: Academic Press.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125–144.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgment of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology, 56A*, 865–890.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58*, 313–323.

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

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 837–854.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition, 18*, 537–545.

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

Cheng, P. W., & Holyoak, K. J. (1995). Complex adaptative systems as intuitive statisticians: Causality, contingency and prediction. In H. L. Roitblat & J.-A. Meyer (Eds.), *Comparative approaches to cognitive science* (pp. 271–302). Cambridge, MA: MIT Press.

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

Cobos, P. L., Caño, A., López, F. J., Luque, J. L., & Almaraz, J. (2000). Does the type of judgement required modulate cue competition? *Quarterly Journal of Experimental Psychology, 53B*, 193–207.

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

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin, 8*, 214–220.

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

De Houwer, J. (2002). Forward blocking depends on retrospective inferences about the presence of the blocked cue during the elemental phase. *Memory & Cognition, 30*, 24–33.

De Houwer, J., & Beckers, T. (2002). Higher-order retrospective revaluation in human causal learning. *Quarterly Journal of Experimental Psychology, 55B*, 137–151.

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology, 55A*, 965–985.

De Houwer, J., Beckers, T., & Vandorpe, S. (2005). Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learning & Behavior, 33*, 239–249.

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

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology, 49B*, 60–80.

Dickinson, A., Shanks, D. R., & Evenden, J. (1984). Judgement of act–outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, 36A*, 29–50.

Frankel, M. S., & Siang, S. (1999). *Ethical and legal aspects of human subjects research on the Internet.* [Report of a workshop convened by the American Association for the Advancement of Science, Program on Scientific Freedom, Responsibility, and Law, Washington, DC.] Retrieved May 9, 2005, from http://www.aaas.org/spp/dspp/sfrl/projects/intres/main.htm

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93–104.

Gredebäck, G., Winman, A., & Juslin, P. (2000). Rational assessments of covariation and causality. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 190–195). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Hetherington, P. A., & Seidenberg, M. S. (1989). Is there "catastrophic interference" in connectionist networks? *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 26–33). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence, 46*, 47–75.

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

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of board of scientific affairs' advisory group on the conduct of research on the Internet. *American Psychologist, 59*, 105–117.

Le Pelley, M. E., Oakeshott, S. M., & McLaren, I. P. L. (2005). Blocking and unblocking in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes, 31*, 56–70.

Lewandowsky, S. (1991). Gradual unlearning and catastrophic interference: A comparison of distributed architectures. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data:*

Essays on human memory in honor of Bennet B. Murdock (pp. 445–476). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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, & Cognition, 24*, 672–694.

Lovibond, P. F., Been, S.-L., Mitchell, C. J., Bouton, M. E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition, 31*, 133–142.

Matute, H., Arcediano, F., & Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 182–196.

Matute, H., & Vadillo, M. A. (2005). *Inferring causality and making predictions: Some misconceptions in the animal and human learning literature.* [Paper presented at the Online Conference on Causality organized by the Institute for Cognitive Sciences at Lyon and the Université de Genève.] Retrieved July 21, 2005, from http://www.interdisciplines.org/causality/papers/16

Matute, H., Vegas, S., & De Marez, P. J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 714–725.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–165). San Diego, CA: Academic Press.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11*, 502–506.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press.

Pineño, O., Denniston, J. C., Beckers, T., Matute, H., & Miller, R. R. (2005). Contrasting predictive and causal values of predictors and causes. *Learning & Behavior, 33*, 184–196.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review, 97*, 285–308.

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 research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

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

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 229–261). San Diego, CA: Academic Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*, 165–173.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1–74.

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

Vadillo, M. A., Miller, R. R., & Matute, H. M. (2005). Causal and predictive-value judgments, but not predictions, are based on cue–outcome contingency. *Learning & Behavior, 33*, 172–183.

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.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*, 171–180.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 53–76.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*, 222–236.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19*, 231–241.

Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology, 103*, 284–297.

Wasserman, E. A. (1990). Attribution of causality to common and distinctive elements of compound stimuli. *Psychological Science, 1*, 298–302.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response–outcome contingencies under free-operant procedures. *Learning and Motivation, 14*, 406–432.

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

White, P. A. (2003). Effects of wording and stimulus format on the use of contingency information in causal judgment. *Memory and Cognition, 31*, 231–242.