Causal and predictive-value judgments, but not predictions, are based on cue–outcome contingency

MIGUEL A. VADILLO
Universidad de Deusto, Bilbao, Spain

RALPH R. MILLER
State University of New York, Binghamton, New York

and

HELENA MATUTE
Universidad de Deusto, Bilbao, Spain

In three experiments, we show that people respond differently when they make predictions as opposed to when they are asked to estimate the causal or the predictive value of cues: Their response to each of those three questions is based on different sets of information. More specifically, we show that prediction judgments depend on the probability of the outcome given the cue, whereas causal and predictive-value judgments depend on the *cue–outcome contingency*. Although these results might seem problematic for most associative models in their present form, they can be explained by explicitly assuming the existence of postacquisition processes that modulate participants' responses in a flexible way.

Following the work of David Hume (1739/1964), many scholars have viewed the process of causal induction as the progressive strengthening of an association between the mental representation of the cause and the mental representation of the effect. This idea has been at the basis of the associative models of causal learning (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994). Although associative models of learning were initially developed to explain the data obtained in animal conditioning experiments (e.g., Rescorla & Wagner, 1972), the observed similarities between animal conditioning and human causal learning (Alloy & Abramson, 1979; Dickinson, Shanks, & Evenden, 1984) suggested that associative models could also be used to explain causal judgments. According to these models, whenever two events appear close together in time, the strength of the association between their mental representations grows. When people are asked to assess whether there is a causal relation between two events, they presumably base their estimations on the strength of the association between the representation of the potential cause and the representation of the effect.

Associative models have been quite successful in explaining some of the most relevant phenomena observed in causal induction experiments and have also helped to establish a link between research in human cognition and research in animal learning. However, one of the main limitations of these models is that they assume that organisms store and use a single index of the relationship between events (i.e., the associative strength). This means that most associative models cannot make a distinction between the different indexes that people can use to describe the way in which events are related. These models would predict that participants asked to rate the degree of relationship between two events would give the same estimation regardless of the specific question they are asked. That is, according to associative models, the response of participants who are asked to predict the occurrence of one of the events will be similar to that of participants who are asked to rate the strength of the causal relation between both events. Not only theoretical models, but researchers in general, have tended to use terms such as *causal induction*, *predictive value* (of a cue), and *prediction* (of the outcome) as synonymous.

Under certain circumstances, however, one could expect different responses to different types of questions. Imagine a situation in which we have to assess the relationship between a number of patients taking a medicine and their developing an allergic reaction as a secondary effect of the medicine. The allergy appears in 75% of the patients, no matter whether they have taken the medicine

or not. If we were asked to judge the causal relation between the medicine and the allergy, we should conclude that there is no such relationship because the outcome (O) is equally probable both in the presence and in the absence of the cue (C) [i.e., $P(\text{O}|\text{C}) = P(\text{O}|\sim\text{C})$]. Similarly, if we were asked to rate the degree to which taking the medicine is a good predictor of the development of the allergic reaction, we should conclude that taking the medicine is not a good predictor because the allergy occurs anyway. However, if we were asked to predict how likely it is that a patient who has taken the medicine will develop the allergic reaction, we should strongly predict the development of such allergy because the probability of the outcome given the cue is high [i.e., $P(\text{O}|\text{C}) = .75$].[1] Thus, whereas participants asked to predict the outcome given the presence of the cue should base their judgments on the probability of the outcome given the cue, $P(\text{O}|\text{C})$, participants asked to estimate the strength of the predictive or causal relation (i.e., the predictive or causal value of the cue) should instead look for differences in the probability of the outcome in the presence and in the absence of the cue, $P(\text{O}|\text{C}) - P(\text{O}|\sim\text{C})$, a statistical measure known as ΔP (Allan, 1980; Cheng & Novick, 1990, 1992). So we should expect very different judgments depending on whether participants are asked to predict whether the outcome will occur or to estimate the causal or the predictive value of the cue. However, if participants based their judgments, regardless of their being predictions or causal estimations, directly on the output of an associative mechanism such as that proposed by most associative models (e.g., Dickinson & Burke, 1996; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994), participants would not be able to make such a distinction. Predictions, causal judgments, and predictive-value judgments would reflect the strength of the association between the cause and the effect.

Existing data suggest that people are indeed able to make at least some of these subtle distinctions. For example, Wasserman, Elek, Chatlosh, and Baker (1993, Experiment 3), asked their participants to rate not only the causal strength of the relationship between a cue and an outcome, but also the probabilities of the outcome both in the presence and in the absence of the cue. Although they did not directly compare the causal and the probability ratings, their results show that participants' causal ratings correlated with ΔP, whereas ratings of the probability of the outcome in the presence of the cue correlated with the objective $P(\text{O}|\text{C})$. This indicates that participants were clearly able to distinguish between these types of judgments and gave different responses in each case.

Other studies have explicitly compared ratings to different types of test questions and have shown that people's judgments vary depending on the type of question used to request their judgments. In a study by Crocker (1982), participants had to assess the relationship between exercising the day before a match and winning the match. When they were asked to rate the relationship between exercising and winning, judgments were more in-

fluenced by the number of trials in which the cue (exercising the day before) and outcome (winning) were present. However, when participants were asked to rate the relationship between exercising and losing, judgments were more influenced by the trials on which exercising occurred but the hypothetical individual lost. That is, the weight given to each trial type depended on the type of question used. Recently, White (2003) has provided a more systematic study of the effect of the question wording on the weighting of the different trial types. He found that subtle manipulations, like asking for the effect of the cause in a negative way (i.e., asking for the effect of withdrawing the potential cause instead of asking for the effect of introducing it) or using a passive grammatical structure (i.e., asking to what degree the effect is caused by the potential cause) resulted in participants giving a different weight to each trial type.

Matute, Arcediano, and Miller (1996), using a relative stimulus validity design, also observed that participants' judgments changed depending on the type of question. In this procedure, participants are exposed to two compounds of stimuli, AX and BX. For one group of participants, one of these compounds is always followed by an effect [$P(\text{O}|\text{AX}) = 1$], while the other is always followed by its absence [$P(\text{O}|\text{BX}) = 0$]. For a different group of participants, both compounds are followed by the effect on 50% of the trials [$P(\text{O}|\text{AX}) = P(\text{O}|\text{BX}) = .50$]. Thus, Cue X, which is the target cue, is followed by the outcome in 50% of the trials in both groups. Matute et al. (1996) reported that when asked to rate the causal relation between the target cue, X, and the effect, participants exposed to the first condition gave lower ratings than participants exposed to the second one. That is, the presence of a strong potential cause, A, made participants disregard X as a causal factor. This replicates the well-known results observed with nonhuman animals under the same design (e.g., Wagner, Logan, Haberlandt, & Price, 1968; Wasserman, 1974). However, when asked to judge the degree of co-occurrence between X and the outcome (i.e., the percentage of times Cue X had been followed by the outcome), judgments did not differ between the groups. This means that people recognize that some events are sometimes followed by others, even when no causal relation is perceived between them. That is, contrary to what associative models pose, people can give different ratings depending on what they are being asked (but see Cobos, Caño, López, Luque, & Almaraz, 2000, for a different result).

Similarly, in another set of experiments, Matute, Vegas, and De Marez (2002) also observed differences between participants' causal and prediction judgments. They exposed their participants to an acquisition-extinction paradigm in which a cue was present in all trials. It was always followed by an outcome during the first half of treatment (acquisition) and during the last half of treatment was never followed by the outcome (extinction). On each trial, half of the participants in each group were asked to predict the likelihood that the outcome would occur. The remaining participants were asked to rate the

causal relation between the cue and the outcome. The result was that participants asked to predict the outcome were more sensitive to the most recent information that they received (i.e., extinction trials) than participants asked to assess the causal relation.

As in the case of Matute et al. (2002), in the present research, we focus on the different responses people give to different questions. But instead of assessing differences due to the temporal distribution of the information, we tested whether participants' responses adjust to the differences between prediction and causation mentioned above even when the *cue–outcome contingency* remains unchanged during training. Specifically, we tested whether participants' judgments when they are asked *to predict the likelihood that the outcome will occur* are by default more sensitive to the probability of the outcome in the presence of the cue [i.e., $P(O|C)$], whereas *causal judgments* are more sensitive to the overall contingency between the cue and the outcome (i.e., ΔP). In addition, we examined responses to another type of predictive question in which participants were asked to rate the *predictive value* (or predictiveness) of the cue. Although this latter question is also focused on the predictive relation between the cue and the outcome, it does not require participants to make a prediction concerning the outcome. Instead, it requires them to estimate the value of the cue as a predictor of the outcome. In other words, it encourages participants to take into account not only how likely the outcome is in the presence of the cue [i.e., $P(O|C)$] but also how likely it is in its absence [i.e., $P(O|{\sim}C)$]. Indeed, a cue can be a good predictor of an outcome only if it adds some information regarding the probability of the outcome's occurrence. That is, the probability of the outcome's occurrence has to be different in the presence than in the absence of the cue for the cue to have predictive value. This concept of predictive value is similar to what animal researchers usually mean when describing the conditioned stimulus–unconditioned stimulus relationship in terms of predictiveness (Rescorla, 1968). On the other hand, the main difference between making a prediction and estimating the predictive value (or predictiveness) of a cue is that in the first case the participant is rating his/her expectancy that the outcome will follow the cue, whereas in the second case the participant is assessing whether the cue is a reliable predictor (or indicator) that the outcome will occur. Thus, we expected that judgments in reply to the predictive-value question would be closer to causal judgments than to prediction judgments.

The distinction between different types of predictive judgments must be emphasized because they are often confused in the literature, and many experiments have used different types of predictive questions. Some researchers dealing with predictive judgments have asked participants to predict the outcome on the basis of the presence of the cue (or to rate their expectancy of the outcome), which is equivalent to what we are here calling prediction judgment (e.g., De Houwer, Beckers, &

Glautier, 2002; Livesey & Boakes, 2004; Matute et al., 2002; Vadillo, Vegas, & Matute, 2004). By contrast, other researchers have asked participants to rate the degree to which the cue is a predictor (or an indicator) of the outcome, which is equivalent to what we are here calling predictive-value judgments (e.g., Cobos, López, Caño, Almaraz, & Shanks, 2002; Matute et al., 1996; Pineño, Denniston, Beckers, Matute, & Miller, 2005). As argued above, each type of predictive judgment supports a different normative analysis, which means that they should not be used interchangeably. The empirical question of whether participants actually perceive differences between them was tested in the present series of experiments.

EXPERIMENT 1

In Experiment 1, we exposed participants to a sequence of trials in which the *cue–outcome contingency* was zero, but the probability of the outcome was high both in the presence and in the absence of the cue [i.e., $P(O|C) = P(O|{\sim}C) = .80$, $\Delta P = 0$]. Then we assessed their answers to three different questions. The prediction question asked participants to predict to the degree to which they thought that the outcome would occur provided the presence of the cue. The predictive-value question asked them to assess to the degree to which they thought the cue was a good predictor of the outcome. And the causal question asked them to rate to the degree to which they thought that the cue was the cause of the outcome. We expected high ratings for the prediction question because the probability of the outcome given the cue was high. In addition, we expected participants' responses to predictive-value and causal questions to be low, given that a cue is not likely the cause of an outcome, or even a good predictor, if the outcome is equally probable both in its presence and in its absence.

Method

Participants and Apparatus

An important feature of Experiments 1 and 2 is that they were run over the Internet. Web-delivered experiments are not yet common in our research area. However, their advantages and disadvantages, as well as their reliability, have been well studied (for reviews see Birnbaum, 2000; Gosling, Vazire, Srivastava, & John, 2004; Reips & Bosnjak, 2001) and the Internet has been successfully used in experiments concerning probability learning, causal induction, and decision making (e.g., Birnbaum, 1999; Birnbaum & Wakcher, 2002; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). In order to control for the potential noise that the use of this methodology could have introduced in Experiments 1 and 2, Experiment 3 was aimed to replicate and extend the results of these experiments in the laboratory.

One hundred and sixty-seven anonymous Internet users who visited our virtual laboratory (http://www.labpsico.com) volunteered to participate in the experiment. They were tested using a computerized version of the allergy task frequently used in human causal learning experiments (e.g., Wasserman, 1990). The experimental program was written in JavaScript, so that it could work on common Internet browsers.[2] The participants were not asked to provide any personal data: All of them were anonymous and voluntary, accord-

ing to ethical standards for human research over the Internet (Frankel & Siang, 1999).

Design and Procedure

The experiment used a within-subjects design in which all participants received the three types of questions under study at the end of training. During training, participants were shown the records of 60 fictitious patients, one per trial. In each trial, participants saw first whether or not that trial's patient had taken a medicine called Dugetil. In order to keep their attention, participants were next asked to predict the outcome (an allergic reaction) or its absence by giving only a yes or no response. After participants had entered their response, they were told whether that patient had actually developed the allergic reaction and were then allowed to see the medical record of the next patient, and so on. These yes/no responses are frequently used in human contingency learning experiments as a means of retaining participants' attention and the available evidence shows that, although requesting numerical judgments in all trials does affect final judgments, these binary responses have no effect on the final judgment given at test (Matute et al., 2002).

There were four types of trials: trials in which the patient had taken Dugetil and developed the allergic reaction (Trial Type A), trials in which the patient had taken Dugetil but not developed the allergic reaction (Trial Type B), trials in which the patient had not taken Dugetil but developed the allergic reaction (Trial Type C), and trials in which the patient had not taken Dugetil nor developed the allergic reaction (Trial Type D). The frequency of each of these trial types was 24, 6, 24, and 6, respectively. This yielded probabilities of the outcome (the patient developing the allergic reaction) given either the presence or the absence of the cue (the patient taking or not taking Dugetil) equal to .80. Trials were ordered in a random sequence.

Before the 60 training trials, participants were given a screen of instructions. A translation from Spanish of these instructions reads as follows:

> This experiment studies people's ability to learn relations between different events. Imagine that you are a specialist who wants to study the degree to which 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.

Once they had seen the 60 patients' records, a screen appeared in which there were three questions and three scales upon which participants were asked to enter their judgments by clicking one point on the scale. Although the scale was not itself labeled with numbers, when participants moved the mouse over some part of the scale, a text box appeared indicating the numerical value of that point, ranging from 0 to 100. One of the questions, the causal one, asked them to assess the causal relation between the medicine and the allergic reaction. The translation from Spanish of this question is: *To what extent do you think that Dugetil is the cause of the allergic reaction?* The opposite ends of the scale below this question were labeled as *It is definitely not the cause* and *It is definitely the cause*. A second question, the predictive-value question, asked them to rate the degree to which the medicine was a good a predictor of the allergic reaction. The translation from Spanish is as follows: *To what extent do you think that taking Dugetil is a good predictor of the allergic reaction?* The end points of the scale below this question were labeled as *It is not a reliable predictor* and *It is a reliable predictor*. Finally, the prediction question asked participants to predict the extent to which they would expect a patient that has taken the medicine to develop the allergic reaction. A translation from Spanish is as follows: *If a patient has taken Dugetil, to what extent do you think that this patient will develop the allergic reaction?* The end points of the scale below this question were labeled as *Definitely will not develop it* and *Definitely will develop it*. The three questions were presented on the same screen, but their position was counterbalanced across participants. The participants were asked to answer these questions from top to bottom, but were allowed to change their responses to each question before saving the data.

Results and Discussion

The mean judgments given at test for each question are depicted in Figure 1. As expected, these means show that the response to the prediction question was considerably higher than those given to either the causal or predictive-value questions and that there are no remarkable differences between these latter ones. These impressions were confirmed by a one-way repeated measures analysis of variance (ANOVA) performed on the judgments at test, which revealed an effect of the type of question [$F(2,332) =$ 23.61, $MS_e = 445.6$, $p < .001$], and by planned comparisons (with α divided by the number of contrasts, accord-
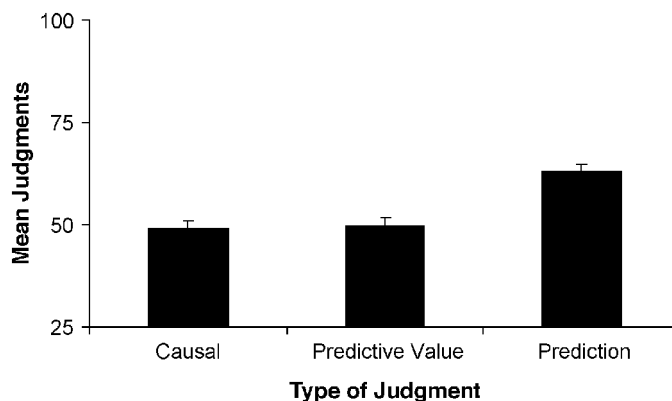

Figure 1. Mean judgments in response to three types of test question (causal, predictive-value, and prediction) in Experiment 1, in which $P(O|C) = P(O|\sim C) = .80$. Error bars represent the standard error of each mean.

ing to the Bonferroni correction procedure, $\alpha = 0.05/3 = 0.017$), which revealed significant differences between prediction judgments and both causal [$t(166) = 6.83$, $p < .001$] and predictive-value judgments [$t(166) = 5.35, p < .001$], whereas no differences were found between causal and predictive-value judgments [$t(166) = 0.2, p = .84$]. The results of Experiment 1, taken together with previously reported experiments (Matute et al., 1996; Matute et al., 2002), provide support for the proposal that participants are sensitive to the type of question used to assess their judgments about the relationship between a cue and an outcome.

EXPERIMENT 2

The results of Experiment 1 clearly show that people can distinguish between making a prediction and assessing the strength of a causal or a predictive relation, although different mechanisms could be proposed to account for the pattern of results. On the one hand, the fact that prediction judgments tend to be higher than causal and predictive-value judgments when $P(O|C)$ is high but ΔP is low suggests that, as expected, prediction judgments are sensitive to $P(O|C)$ while causal and predictive-value judgments are sensitive to ΔP. On the other hand, our results could reflect participants' giving less weight to $P(O|\sim C)$ when making predictions than when assessing predictive value or causal relations. Moreover, it could simply be that people are more conservative when considering the causal strength of a cue–outcome relation than when making predictions, an idea that some could interpret as consistent with Pineño et al.'s (2005) proposal that causal judgments are more sensitive to stimulus competition than are predictive-value judgments.[3]

However, both causal and predictive-value judgments were considerably higher than 0 in Experiment 1, contrary to what we would expect if these judgments were based on ΔP. This result might have been due to the use of an inappropriate scale for registering participants' judgments. Although a scale labeled from 0 to 100 is suitable for prediction judgments, it might be inappropriate for causal and predictive-value judgments, given that their normative referent, ΔP, can yield negative results. In this situation, participants might have interpreted the lower part of the scale (i.e., from 0 to 50) as representing the negative contingency. If so, they might have given an estimation of 50 because they wanted to indicate the absence of a causal relation or a null predictive value. Thus, it might be better to assess people's ability to differentiate prediction, predictive-value, and causal questions with a procedure that does not depend so much on their interpretation of the scale. The fact that the scale for each judgment was labeled in a different way is another possible source of the differences observed in Experiment 1, which makes it even more desirable that we look for alternative strategies. Because of this, in Experiment 2, instead of looking for absolute differences in the ratings to each question, we looked for

manipulations that could affect one type of judgment without having an effect on the other ones. For example, varying ΔP between groups while keeping $P(O|C)$ constant should have an effect on predictive-value and causal judgments, but no effect on prediction judgments. We tested this prediction in Experiment 2.

Alternatively, the high ratings observed in prediction judgments of Experiment 1 could have been due to a preasymptotic bias in participants' responses. It is well known that judgments tend to be preasymptotically high even under null contingencies when the probability of the outcome is high (Allan & Jenkins, 1983; Chatlosh, Neunaber, & Wasserman, 1985; Shanks, 1987; Shanks, López, Darby, & Dickinson, 1996; for an analogous effect in animal conditioning, see Benedict & Ayres, 1972; Kremer & Kamin, 1971). Moreover, this preasymptotic outcome-density bias is a straightforward prediction of the Rescorla–Wagner model (Allan, 2003; Shanks, 1995). However, although this bias can explain why judgments greater than 0 were observed under a null *cue–outcome contingency*, it cannot account for the observed differences between the different types of judgments (i.e., according to this explanation, all judgments should have been equally high). Nevertheless, in Experiment 2 we decided to minimize the potential influence of the preasymptotic outcome-density bias by doubling the number of trials.

Method

Participants and Apparatus

Sixty-nine anonymous participants volunteered for the experiment. All of them performed the experiment through the Internet, as in Experiment 1. The computer program randomly distributed the participants across two experimental groups, which resulted in 35 participants in Group 0.80-0.80 and 34 participants in Group 0.80-0.10.

Design and Procedure

The participants in Group 0.80-0.80 were exposed to a sequence of 120 trials with a null contingency but with a high probability of the outcome both in the presence and in the absence of the cue [i.e., $P(O|C) = P(O|\sim C) = .80$, $\Delta P = 0$]. The frequencies of each trial type, A, B, C, and D, for this group were 48, 12, 48, and 12, respectively. The participants in Group 0.80-0.10 were exposed to a different sequence of 120 trials with a high contingency and a high probability of the outcome in the presence of the cue [i.e., $P(O|C) = .80$, $P(O|\sim C) = .10$, $\Delta P = .70$]. The frequencies of each trial type, A, B, C, and D, for this group were 48, 12, 6, and 54, respectively. All other procedural details were kept the same as in Experiment 1. According to our hypothesis, prediction judgments should be similar in both groups, because the $P(O|C)$ is .80 in both cases. However, the predictive value and causal value of the cue should be higher in Group 0.80-0.10 than in Group 0.80-0.80, because the contingency is higher in that group.

Results and Discussion

Figure 2 shows the mean judgments given for each type of test question in Experiment 2. As can be seen in this figure, the results for Group 0.80-0.80 replicated those obtained in Experiment 1: Responses to the prediction question were considerably higher than responses to the causal and predictive-value questions. In addition,
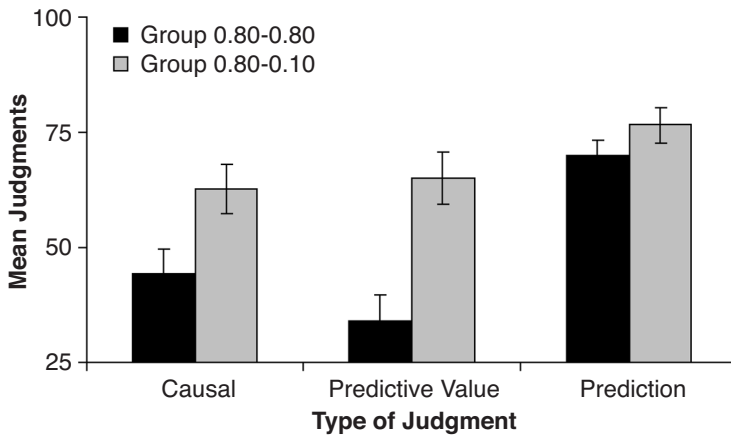
Figure 2. Mean judgments in response to three types of test question (causal, predictive-value, and prediction) for the two groups in Experiment 2. Error bars represent the standard error of each mean.

participants in Group 0.80-0.10 gave, relative to the participants in Group 0.80-0.80, equally high ratings for the prediction question but higher ratings for the causal and predictive-value questions. A 3 (type of question: causal, predictive-value, and prediction) $\times$ 2 (group: 0.80-0.80 vs. 0.80-0.10) ANOVA yielded main effects of both type of question [$F(2,134) = 23.17$, $MS_e = 485.51$, $p < .001$] and group [$F(1,67) = 11.4$, $MS_e = 1,564.11$, $p < .01$]. Moreover, an interaction was found between these two variables [$F(2,134) = 5.28$, $MS_e = 485.51$, $p < .01$], indicating that the type of test question used had different effects on the two groups. As we expected, planned comparisons (with α adjusted according to the Bonferroni correction procedure, $\alpha = 0.05/5 = 0.01$) showed that there were no significant differences between the prediction judgments given by participants in Group 0.80-0.80 and participants in Group 0.80-0.10 [$t(67) = 1.28$, $p = .21$], but there were differences between the two groups in their responses to the predictive-value question [$t(67) = 3.94$, $p < .001$]. The expected difference in the causal judgments is also evident in Figure 2 [$t(67) = 2.37$, $p = .02$], though it is only marginally significant in light of the Bonferroni correction procedure. Planned comparisons also showed that, as expected, in Group 0.80-0.80 responses to the prediction question were higher than responses to the causal question [$t(34) = 4.63$, $p < .001$], and higher than responses to the predictive-value question [$t(34) = 7.09$, $p < .001$].

The results of this experiment replicate those of Experiment 1; that is, when there is a large difference between $P(O \mid C)$ and ΔP, judgments to each type of question diverge. In addition, the general pattern of results provides stronger support for our hypothesis, because it shows that there are manipulations, like varying $P(O \mid \sim C)$, which affect responses to some of the questions but not others. Specifically, this manipulation affects predictive-value judgments, but not prediction ones. The question of whether or not this manipulation affects causal judg-

ments remains unclear in this experiment, given that the difference between causal ratings in both groups is marginally significant if the Bonferroni correction procedure is applied. Thus, the prediction that causal judgments should vary depending on ΔP was tested again in Experiment 3.

One reason for conducting Experiment 2 was that the scale used in Experiment 1 to request causal and predictive-value judgments might not have been appropriate for assessing participants' subjective ratings, because the parameter assumed to be normative for causal and predictive-value judgments, ΔP, can adopt negative values that did not appear in the scale we used, and also because each type of question had its corresponding scale labeled in a different way. However, the results obtained in Experiment 2 show that there are manipulations that can affect predictive-value and perhaps causal judgments without having an effect on prediction judgments, a finding that is independent of the scale used to request these judgments and that overcomes this potential limitation of Experiment 1.

EXPERIMENT 3

The preceding experiments demonstrated not only that participants respond differently to causal, predictive-value, and prediction questions, but also that there are manipulations that influence responses to some of these questions while having no effect on others. Specifically, manipulations of contingency seem to have little or no effect on responses to the prediction question if the probability of the outcome given the cue is kept the same across conditions. However, this manipulation has a clear effect on responses to causal and predictive-value questions. What remains to be tested is the complementary prediction—that is, that there can be manipulations that influence prediction judgments without having an effect on causal and predictive-value judgments. According to

our hypothesis, manipulating the probability of the outcome given the cue without changing the contingency should yield this pattern of results. This prediction was tested in Experiment 3.

In addition, the data from Experiments 1 and 2 were collected using an Internet-based methodology, which poses the question of whether these results might not be replicable under more controlled situations, like those of the traditional laboratory. In order to address this question, Experiment 3 was performed in the laboratory and it included two conditions that tested the same hypothesis as Experiments 1 and 2, although using slightly different contingencies.

Method

Participants and Apparatus

Ninety-nine undergraduate students from Deusto University volunteered to take part in the experiment. Random assignment of participants resulted in 35 participants in Group 0.50-0.00, 32 participants in Group 1.00-1.00, and 32 participants in Group 1.00-0.50. The participants performed the experiment in a large computer room. Adjacent participants were seated at about 1.5 m apart from each other and were exposed to different experimental conditions. Although the experiment was not performed over the Internet, the experimental program was an adapted version of the same JavaScript program used in Experiments 1 and 2.

Design and Procedure

Three groups of participants were exposed to 120 trials that reflected particular cue–outcome contingencies and probabilities of the outcome given the cue. Group 0.50-0.00 was exposed to an intermediate probability of the outcome given the cue and an intermediate *cue–outcome contingency* [i.e., $P(O|C) = .50$, $P(O|\sim C) = .00$, $\Delta P = .50$]; the frequencies of each trial type, A, B, C, and D, for this group were 30, 30, 0, and 60, respectively. Group 1.00-1.00 was exposed to a high probability of the outcome given the cue but a null *cue–outcome contingency* [i.e., $P(O|C) = P(O|\sim C) = 1.00$, $\Delta P = .00$]; the frequencies of each trial type, A, B, C, and D, for this group were 60, 0, 60, and 0, respectively. Finally, Group 1.00-0.50

was exposed to a high probability of the outcome given the cue and a medium *cue–outcome contingency* [i.e., $P(O|C) = 1.00$, $P(O|\sim C) = .50$, $\Delta P = .50$]; the frequencies of each trial type, A, B, C, and D, for this group were 60, 0, 30, and 30, respectively. All other procedural details were kept the same as in Experiments 1 and 2. According to our hypothesis, Groups 1.00-1.00 and 1.00-0.50 should differ in their responses to the causal and predictive-value questions but not in their response to the prediction question (a result that would replicate the pattern of results observed in the previous experiments). Moreover, Groups 0.50-0.00 and 1.00-0.50 should differ in their response to the prediction question but not in their response to the causal and predictive-value questions.

Results and Discussion

The mean responses to each type of test question can be seen in Figure 3. A 3 (type of question: causal, predictive-value, prediction) × 3 (group: 0.50-0.00, 1.00-1.00, and 1.00-0.50) ANOVA performed on participants' judgments yielded a main effect for both type of question [$F(2,192) = 25.15$, $MS_e = 518.24$, $p < .001$] and group [$F(2,96) = 9.75$, $MS_e = 1,420.3$, $p < .001$]. More interesting was that an interaction between these factors was also observed [$F(4,192) = 7.06$, $MS_e = 518.24$, $p < .001$], indicating that the effect of the type of question depended on the group.

The general pattern of judgments given by participants in Group 1.00-1.00 supports the idea that prediction judgments tend to be higher than causal and predictive-value judgments when the $P(O|C)$ is high but ΔP is low. This impression was confirmed by planned comparisons (with α corrected, according to the Bonferroni correction procedure, $\alpha = 0.05/9 = 0.0055$), which showed that prediction judgments were higher than either causal or predictive-value judgments [$t(31) = 4.84$, $p < .001$, and $t(31) = 5.62$, $p < .001$, respectively]. Moreover, the difference between causal and predictive-value judgments was nonsignificant [$t(31) = 0.46$, $p = .65$]. These
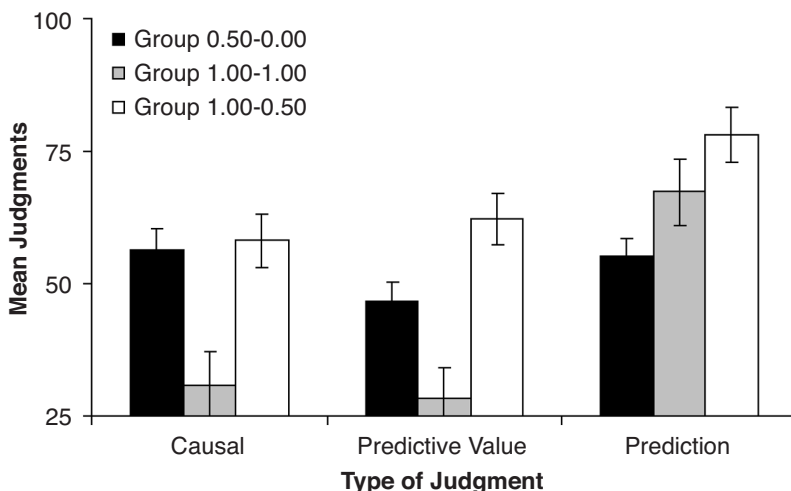
Figure 3. Mean judgments in response to three types of test question (causal, predictive-value, and prediction) for the three groups in Experiment 3. Error bars represent the standard error of each mean.

results replicate those found in Experiment 1 as well as in Group 0.80-0.80 in Experiment 2. This indicates that the results of these Internet-based experiments can be replicated in the laboratory.

In addition, planned comparisons showed that, as expected, responses to both causal and predictive-value questions were lower in Group 1.00-1.00 than in Group 1.00-0.50 [$t(62) = 3.39$, $p < .0055$, and $t(62) = 4.47$, $p < .001$, respectively], and that the response to the prediction question did not differ significantly between groups [$t(62) = 1.3$, $p = .20$]. These results replicate those of Experiment 2 regarding the differential effect of manipulating $P(O|{\sim}C)$ on predictive-value and prediction judgments. That is, when the probability of the outcome given the cue is kept constant, manipulations of contingency had an effect on predictive-value judgments but not on prediction judgments. Importantly, the fact that the difference in causal judgments between the two groups with the same $P(O|C)$ but different ΔP was now significant confirms the tendency that was already apparent in Experiment 2. The present experiment was conducted under controlled laboratory conditions, which suggests that the marginally significant effect observed in Experiment 2 was probably due to a lack of statistical power, perhaps induced by the larger variance usually observed in Internet-based experiments.

Contrary to our hypothesis, prediction judgments tended to be somewhat higher in Group 1.00-0.50 than in Group 1.00-1.00. This tendency could indicate that prediction judgments are actually influenced by $P(O|{\sim}C)$, although in a lesser degree than causal and predictive-value judgments (a result that could be seen as supporting the differential sensitivity view proposed by Pineño et al., 2005; but see note 3). However, the fact that this difference was nonsignificant in both Experiments 2 and 3 suggests that this difference is not robust.

In addition, the comparison of Groups 0.50-0.00 and 1.00-0.50 indicates that responses to some questions were affected by a change in the $P(O|C)$ that left the *cue–outcome contingency* unchanged, while others remained the same. As expected, planned comparisons showed that responses to the causal question were not significantly different between these two groups [$t(65) = 0.29$, $p = .77$], and that responses to the prediction question differed [$t(65) = 3.75$, $p < .001$]. The difference in predictive-value judgments, which is contrary to our hypothesis and to the pattern of results observed in the previous experiments, is not significant according the Bonferroni correction [$t(65) = 2.59$, $p = .012$], but admittedly there is a tendency toward a difference. It is not clear to us whether this is due to a true absence of difference or to a lack of statistical power, given that this is the first time that we have observed this tendency. Because of this, we prefer to draw no firm conclusions about the predictive-value question in this experiment at this time. Nevertheless, what is clear is that the results of the causal and prediction questions still confirm the general pattern of results of the previous experiments. That is, when the probability of the outcome given the cue is manipulated but *cue–outcome contingency* is kept constant, responses to the prediction question vary but responses to the causal question do not. This general pattern of results is evident both in the laboratory and in the Internet studies.

GENERAL DISCUSSION

Our general pattern of results is consistent with the view that people's judgments about the relationship between a cue and an outcome can be affected by the type of question used to assess these judgments (Crocker, 1982; Matute et al., 1996; Matute et al., 2002; Pineño et al., 2005; White, 2003). Specifically, when participants are asked to rate how likely it is that an outcome will occur, their responses are sensitive to the objective probability of the outcome given the presence of the cue. However, when they are asked to rate the cue's causal relation with the outcome (and often when they are asked to assess the predictive value of the cue), their responses are more sensitive to the difference between the probability of the outcome in the presence and in the absence of the cue. This interpretation is also consistent with the results of many experimental studies that have asked participants to judge the probability of the outcome given the presence of the cue and have obtained accurate estimations of this probability (e.g., Wasserman et al., 1993).

A major finding here is that participants are able to make distinctions not only between causal and prediction judgments but also between predictive-value and prediction judgments. This means that they can perceive differences between different types of predictive questions, even though all of them refer to a noncausal probabilistic relationship between the cue and the outcome. This result emphasizes the importance of researchers' being careful when using predictive questions to assess participants' judgments, given that subtle differences in the wording of this question can have a great impact on the observed judgments, a prescription that will become more and more important as interest in the comparison between causal and predictive learning grows (e.g., De Houwer et al., 2002; Matute et al., 2002; Pineño et al., 2005).

The results of all these studies taken together pose problems for some of the main associative models of human causal and predictive learning (e.g., Dickinson & Burke, 1996; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994). It is well known that the asymptotic outcome of the associative learning rule proposed by Rescorla and Wagner, which inspired later associative models, reflects the statistical contingency between the cue and the outcome. That is, once the association between the cue and the outcome has reached its asymptote, the value of its associative strength equals the statistical measure of contingency ΔP (Chapman & Robbins, 1990).[4] Thus, the Rescorla–Wagner model could easily account for participants' assessments of causal strength and predictive value, given that they seem to be sensitive to ΔP. Importantly, however, the Rescorla–Wagner model cannot account in its present form for participants' predictions of the outcome occurrence, which cannot be com-

puted by a contingency-based learning algorithm. Below we will suggest ways in which several major theories, including the Rescorla–Wagner model, could be extended to account for the present results.

Statistical Reasoning Account

The results of the three experiments we present here could be easily explained in terms of statistical reasoning processes. Such an explanation would simply assume that during training participants store the information presented (e.g., accumulating the frequencies of each type of trial) and that this information can be used in a flexible way when they are asked about the relationship between the cue and the outcome. Depending on the type of question asked, participants would use this information to compute the statistical index most relevant to answer that question. Although most statistical models have focused mainly on causal (Cheng, 1997; Cheng & Novick, 1990, 1992) and contingency (Allan, 1980) judgments, the basic processes suggested by these models can also underlie prediction judgments. Moreover, if the computation of statistical indexes like ΔP is algorithmically based on any sort of contrast of conditional probabilities, $P(O|C)$ and $P(O\sim|C)$, then these models presuppose the ability to compute the specific conditional probabilities that underlie prediction judgments.

Although this general perspective is supported by our data, some models of causal induction, specifically the power PC theory (Cheng, 1997), might have problems explaining part of the results. According to Cheng, humans intuitively know that the causal relations between events underlie covariation relations between events, but also that covariation and causation are different things. In other words, the strength of the causal relation between a cue and an outcome (i.e., the causal power) can be different from the observed contingency between those events. If this is the case, in some situations different judgments would be expected depending on whether participants are asked to rate the strength of the causal relation or they are asked to give an estimation of the degree of covariation between a cue and an outcome. Specifically, according to the power PC theory, ΔP underestimates the generative causal power (the ability of a cue to cause an effect) as $P(O|\sim C)$ increases. However, this prediction is not supported by the results of Experiment 3. In that experiment, two groups of participants, 0.50-0.00 and 1.00-0.50, were exposed to the same *cue–outcome contingency* but different values of $P(O|\sim C)$, which means that causal judgments should be larger in Group 1.00-0.50 than in Group 0.50-0.00. But our results show that the causal ratings were similar in the two groups. Although these results could be due to the use of a causal question instead of a counterfactual question, which has been shown to be more sensitive to causal power (Buehner, Cheng, & Clifford, 2003), they remain unexplainable by the power PC theory.

Despite this problem of the power PC theory to account for the present results, it is still likely that our participants resolved the task by the use of higher order cognitive processes involving statistical reasoning. In fact, any explanation of these results requires the intervention of complex cognitive processes. Our participants at least had to have been able to understand the different semantics of each type of question and to have used some previously acquired knowledge regarding the differences between assessing causal relationships, assessing predictive values, and making predictions. However, as we shall discuss in the next section, this does not mean that more mechanical, lower level, associative mechanisms need to be excluded from the explanation of these results.

Associative Account

Perhaps the main feature of associative models that prevents them from explaining the present results is their lack of a distinction between learning and performance. In other words, for most of these models (Dickinson & Burke, 1996; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994) the performance of participants in a prediction and in a causal task is solely determined by the strength of the cue–outcome associations. However, many experimental findings do not support this point of view (e.g., Allan, Siegel, & Tangen, 2005; De Houwer et al., 2002; Matute et al., 2002). Nevertheless, there are also some models that suggest a clear distinction between the learning processes and the performance processes, models that could be taken as a starting point for an attempt to provide an associative-based account for the results presented in this paper (e.g., Allan et al., 2005; Miller & Matzel, 1988).

As an example, some ideas underlying the comparator hypothesis could be used with this aim. The comparator hypothesis (Miller & Matzel, 1988; Miller & Schachtman, 1985; see also Denniston, Savastano, & Miller, 2001) was developed in the area of animal conditioning and assumes that learning takes place by means of a noncompetitive, contiguity-based learning algorithm, like the one proposed by Bush and Mosteller (1951). According to Bush and Mosteller, the associative strength between a cue and an outcome is based solely on the contiguity between them and is not affected by the strength of the associations that other cues (including the context) may have to the same outcome. This means that the associative strength of the cue is affected only by the probability of the outcome given the cue and that the probability of the outcome in the absence of the cue has no effect on this associative strength. According to the comparator hypothesis, however, a subsequent comparison process takes place at the performance stage between the associative strength of the target cue and that of all other stimuli that were present along with the target cue during training: Conditioned responses in animals will be observed whenever the associative strength of the target cue is higher than that of the comparator stimuli associated with it.

In its present state, the comparator hypothesis is unable to account for the results presented here. However, some

ideas can be borrowed from it. Of special interest would be the assumptions (1) that learning takes place by a learning algorithm that is sensitive to cue–outcome co-occurrence alone (i.e., akin to Bush & Mosteller's learning rule, instead of the more popular and complex contingency-like learning algorithms; e.g., Rescorla & Wagner, 1972) and (2) that associations might not always be directly expressed in behavior (i.e., that a distinction between learning and performance should be made; see also Allan et al., 2005). From this perspective, accounting for the present results would only require the assumption that participants' predictions are based just on the associative strength of the target cue, computed by a noncompetitive or co-occurrence-based learning algorithm, and that causal, and probably predictive-value judgments as well, are based on a comparison between the associative strength of the target cue and that of the context. When the probability of the outcome is high both in the presence and in the absence of the cue, the learning process leads to a high associative strength for both the cue and the context. Thus, judgments based solely on the associative strength of the cue (prediction judgments) will be high, whereas judgments based on the comparison between the associative strengths of the cue and the context (causal and predictive-value judgments) will be low (Experiment 1). If the probability of the outcome given the cue is kept constant, manipulations of $P(O\,|\sim C)$ will affect only the associative strength of the context. Therefore, the comparison process will yield differences in causal (and probably in predictive-value) judgments but not in prediction judgments (Experiments 2 and 3). In addition, manipulations of $P(O\,|\,C)$ should have a direct effect on prediction judgments, but should not affect causal and predictive-value judgments if ΔP remains unchanged (Experiment 3). Thus, this account can explain qualitatively the results of the three experiments here reported by assuming that the comparator processes can be triggered or not depending on the type of question. Note, however, that this assumption is clearly beyond the scope of the comparator hypothesis in its current state; according to a rigorous interpretation of the model, the comparator processes should always take place, regardless of the type of question used to assess participants' judgments.

In a similar vein, it could also be possible to extend the Rescorla–Wagner model to account for our results by simply assuming that judgments about the cue–outcome relation need not necessarily be based exclusively on the target cue–outcome association. What this extension implies is simply recognizing that the association between the cue and the outcome is not directly mapped onto a response, but that in the process of organizing a response the organism might also use the information contained in other associations. From this point of view, prediction judgments (the only ones which pose problems for the Rescorla–Wagner model) could result from our participants' basing their responses on the addition of the cue–outcome association and the context–outcome association. Although so stated this process might appear strange, the underlying logic is quite simple. In order to make a prediction about an outcome's occurrence, it is necessary to take into account the predictive value of all the cues that are present at a given moment. The predictive value of each cue would be summarized in the strength of its association to the outcome (which would be dependent on ΔP as computed by the Rescorla–Wagner learning algorithm). Thus, the probability of the occurrence of the outcome at any given moment would be the sum of the associative strength of all the cues that are present at that moment.

An important advantage of the use of an associative-based account (either the one based on Bush & Mosteller, 1951, or the one based on the Rescorla–Wagner, 1972, model) is that either one is perfectly compatible with the proposals of many cognitive theories, which are often designed to explain how causal judgments are computed but not how causal information is acquired. Most statistical models assume that causal induction is based on a comparison or contrast of conditional probabilities and that this contrast is guided by top-down processes. However, none of those models states clearly how people acquire the information needed to compute the relevant conditional probabilities, nor do they explain how these conditional probabilities are computed. An implicit assumption is that the conditional probabilities are computed by performing some mathematical operation with the frequencies of events stored in a mental contingency table. If this were the case, causal induction would require a great amount of memory resources and complex induction tasks could easily overload the cognitive system. Part of all these processes could instead be performed by very elemental associative mechanisms that would accomplish the task with little cognitive effort. For example, the noncompetitive, co-occurrence-based associative algorithm proposed by Bush and Mosteller could provide a parsimonious process for computing the conditional probabilities that could then be processed by higher cognitive processes if necessary. Those higher order cognitive processes refer not to how information is acquired, but to how acquired information is processed and used. They speak to *postacquisition reasoning processes* but not to the *learning processes* of causal induction. On the other hand, an associative mechanism alone could perfectly well account for the *learning process*, but not for the *postacquisition reasoning process*. The theoretical interpretation we suggest here tries to keep the best of both associative and statistical models and provides a parsimonious analysis of the processes involved in human causal and predictive judgments.

REFERENCES

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

ALLAN, L. G. (2003). Assessing power PC. *Learning & Behavior*, **31**, 192-204.

ALLAN, L. G., & JENKINS, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning & Motivation*, **14**, 381-405.

ALLAN, L. G., SIEGEL, S., & TANGEN, J. M. (2005) A signal detection analysis of contingency data. *Learning & Behavior*, **33**, 250-263.

ALLOY, L. B., & ABRAMSON, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, **108**, 441-485.

BENEDICT, J. O., & AYRES, J. J. B. (1972). Factors affecting conditioning in the truly random control procedure in the rat. *Journal of Comparative & Physiological Psychology*, **78**, 323-330.

BIRNBAUM, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, **10**, 399-407.

BIRNBAUM, M. H. (ED.) (2000). *Psychological experiments on the Internet.* San Diego: Academic Press.

BIRNBAUM, M. H., & WAKCHER, S. V. (2002). Web-based experiments controlled by JavaScript: An example from probability learning. *Behavior Research Methods, Instruments, & Computers*, **34**, 189-199.

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

BUSH, R. R., & MOSTELLER, F. (1951). A mathematical model for simple learning. *Psychological Review*, **58**, 313-323.

CHAPMAN, G. B., & ROBBINS, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, **18**, 537-545.

CHATLOSH, D. L., NEUNABER, D. J., & WASSERMAN, E. A. (1985). Response-outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning & Motivation*, **16**, 1-34.

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 & Social Psychology*, **58**, 545-567.

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.

COBOS, P. L., LÓPEZ, F. J., CAÑO, A., ALMARAZ, J., & SHANKS, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, **28**, 331-346.

CROCKER, J. (1982). Biased questions in judgment of covariation studies. *Personality & Social Psychology Bulletin*, **8**, 214-220.

DE HOUWER, J., BECKERS, T., & GLAUTIER, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, **55A**, 965-985.

DENNISTON, J. C., SAVASTANO, H. I., & MILLER, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65-117). Hillsdale, NJ: Erlbaum.

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 in cyberspace.* 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.

HUME, D. (1964). *A treatise of human nature* (L. A. Selby-Bigge, Ed.). London: Oxford University Press. (Original work published 1739)

KREMER, E. F., & KAMIN, L. J. (1971). The truly random control procedure: Associative or nonassociative effects in the rats. *Journal of Comparative & Physiological Psychology*, **74**, 203-210.

LIVESEY, E. J., & BOAKES, R. A. (2004). Outcome additivity, elemental processing and blocking in human causality judgements. *Quarterly Journal of Experimental Psychology*, **57B**, 361-379.

MATUTE, H., ARCEDIANO, F., & MILLER, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 182-196.

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, & Cognition*, **28**, 714-725.

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: Academic Press.

MILLER, R. R., & SCHACHTMAN, T. R. (1985). Conditioning context as an associative baseline: Implications for response generation and the nature of conditioned inhibition. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 51-88). Hillsdale, NJ: Erlbaum.

PINEÑO, O., DENNISTON, J. C., BECKERS, T., MATUTE, H., & MILLER, R. R. (2005). Contrasting predictive and causal values of predictors and of causes. *Learning & Behavior*, **33**, 184-196.

REIPS, U.-D., & BOSNJAK, M. (EDS.) (2001). *Dimensions of Internet science.* Lengerich: Pabst.

RESCORLA, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative & Physiological Psychology*, **66**, 1-5.

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. (1987). Acquisition functions in contingency judgment. *Learning & Motivation*, **18**, 147-166.

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

SHANKS, D. R., LÓPEZ, F. J., DARBY, R. J., & DICKINSON, A. (1996). Distinguishing associative and probabilistic theories of human contingency judgment. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 265-311). San Diego: Academic Press.

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., VEGAS, S., & MATUTE, H. (2004). Frequency of judgment as a context-like determinant of predictive judgments. *Memory & Cognition*, **32**, 1065-1075.

VAN HAMME, L. J., & WASSERMAN, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning & 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.

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., 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, & Cognition*, **19**, 174-188.

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

NOTES

1. Of course, the symmetrical rule would apply if we were asked to predict how likely it is that a patient who has not taken the medicine

will develop the allergic reaction. In this case our prediction should just be based on the probability of the outcome in the absence of the cue, $P(O|\sim C) = .75$). For the sake of simplicity, however, we will hereafter refer only to the case in which the participant predicts what will happen when the cue is present.

2. A demonstration of the actual program used can be downloaded from http://paginaspersonales.deusto.es/matute/software.html

3. Note that Pineño et al. (2005) used different terms to describe their questions, but the questions they used were predictive-value questions (as defined herein), not predictions. Thus, the difference that they proposed is between causal and predictive-value judgments. By contrast, the difference that we observed is between prediction judgments and both causal and predictive-value judgments (which did not differ from each other).

4. This property of the Rescorla and Wagner (1972) algorithm requires the assumption that the learning parameter β is equal when the outcome is presented and when it is not.