Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 4, 714–725

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.4.714

Flexible Use of Recent Information in Causal and Predictive Judgments

Helena Matute and Sonia Vegas
Deusto University

Pieter-Jan De Marez
Leuven University

Associative and statistical theories of causal and predictive learning make opposite predictions for situations in which the most recent information contradicts the information provided by older trials (e.g., acquisition followed by extinction). Associative theories predict that people will rely on the most recent information to best adapt their behavior to the changing environment. Statistical theories predict that people will integrate what they have learned in the two phases. The results of this study showed one or the other effect as a function of response mode (trial by trial vs. global), type of question (contiguity, causality, or predictiveness), and postacquisition instructions. That is, participants are able to give either an integrative judgment, or a judgment that relies on recent information as a function of test demands. The authors concluded that any model must allow for flexible use of information once it has been acquired.

Learning to predict the events in our environment is critical for survival. Both humans and other animals are known to learn predictive and causal relations between the events in their environment, and the question of how they do it has preoccupied philosophers and psychologists for many years.

There are many different types of models that have tried to answer this question, and some of them seem to contradict each other. For example, according to statistical or computational models (generally developed in the area of human cognition), when participants are asked to judge the degree to which a potential predictor cue, *C*, predicts a given outcome, *O*, they will take into account all of the information they have and will compute their judgment using a statistical rule such as, for example, ΔP (Allan, 1980). ΔP is computed as the probability of the outcome occurring when *C* is present—that is, $p(O \mid C)$—minus the probability of the outcome occurring when *C* is absent—that is, $p(O \mid \text{no}C)$.

However, according to associative models, generally developed in the area of animal learning, predictive and causal judgments are analogous to the conditioned response that an animal gives in a conditioning experiment. On observing the occurrence of the conditioned stimulus or cue, *C* (e.g., light or tone), the animal predicts the immediate occurrence of the unconditioned stimulus or outcome, *O* (e.g., food or foot shock), and hence, salivates or freezes as a result of this prediction. According to associative theories, the strength of the expectation of the outcome (whether assessed through a conditioned response in rats or through a predictive or causal judgment in humans), will be a function of the strength of the association that is gradually established between *C* and *O* during the acquisition trials (e.g., Allan, 1993; Rescorla & Wagner, 1972). Thus, according to these theories, the strength of the association will be updated on a trial-by-trial basis: Some types of trials will strengthen the association between *C* and *O*, whereas others will decrease it. For example, every time that a *C*–*O* trial occurs, the associative strength will increase, and any time that a *C*–no*O* trial occurs, the associative strength will decrease.

Opposite Predictions of Associative and Statistical Models

Those two sets of models make opposite predictions concerning recency effects (e.g., Chapman, 1991; López, Shanks, Almaraz, & Fernández, 1998). Associative models predict that the final response will differ as a function of the most recent trials, whereas statistical models predict that the order of trials should have no effect on the final response. For example, according to associative theories, a series of *C*–*O* trials followed by a series of *C*–no*O* trials (i.e., acquisition followed by extinction) should lead to a lower judgment than a series in which the *C*–*O* and C–no*O* trials are presented in the reverse order. By contrast, statistical models predict no difference between those two conditions. According to these theories, participants will calculate their final judgment using all the information received during the two stages, and thus, an integrative judgment, rather than a recency-based judgment, should be observed at test.

It is true that the predictions of associative models are parameter dependent (see e.g., Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Wasserman, Kao, Van Hamme, Katagari, & Young, 1996) and that associative models that give less weight to recent trials can easily be developed. Moreover, statistical models could also be developed that compute only the latest trials so that they could best fit the results of extinction experiments and other recency effects (see Miller & Escobar, 2001; Shanks, López,

Helena Matute and Sonia Vegas, Department of Psychology, Deusto University, Bilbao, Spain; Pieter-Jan De Marez, Department of Psychology, Leuven University, Leuven, Belgium.

Support for this research was provided by Department of Education, Universities and Research of the Basque Government Grant PI-2000-12 awarded to Helena Matute. Sonia Vegas was supported by Beca de Formación Investigadores Fellowship BFI00.138 from the Basque Government. Pieter-Jan De Marez was supported by the Socrates undergraduate exchange program of the European Union. We thank Leyre Castro, Andrés Catena, Antonio Maldonado, Nuria Ortega, Oskar Pineño, and Bram Vervliet for insightful discussions concerning these experiments.

Correspondence concerning this article should be addressed to Helena Matute, Departamento de Psicología, Universidad de Deusto, Apartado 1, Bilbao 48080, Spain. E-mail: matute@orion.deusto.es

Darby, & Dickinson, 1996, for discussion). However, the cognitive system may be much more flexible than any of those classes of models predict. As predicted by associative theories, the predictive or causal relations that we once learned may no longer be valid (things change in our environment), and there must be a mechanism that allows us to adapt to these changes. At the same time, however, it might be quite maladaptive to discard our previous knowledge and rely only on the most recent information. Quite possibly, the cognitive system is able to use either the most recent information or all of it as a function of the demands of the environment.

However, the many studies on extinction available in the literature (e.g., Paredes-Olay & Rosas, 1999; Pavlov, 1927) provide clear evidence in favor of the associative prediction and against the statistical prediction. For this reason, our approach in these experiments was to use an extinction design to test whether integrative judgments could also be observed after extinction training as a function of several testing manipulations.

Variables That Might Affect the Observation of One or the Other Result

One of the variables that could affect the observation of trial-order effects is the wording of the test question that is used to request the participants' judgments. These questions generally vary from one published study to another, and the potential influence of this variable is generally overlooked. However, Matute, Arcediano, and Miller (1996) presented evidence that the type of question used to assess participants' judgments affects the results of judgmental experiments. More specifically, they used a cue competition design and observed cue competition effects (i.e., judgments that become biased when there are several cues associated to the same outcome) when participants were asked to give predictive and causal judgments, but not when participants were asked to judge the degree of contiguity (or co-occurrence) between the cue and the outcome (but see also Cobos, Caño, López, Luque, & Almaraz, 2000). Although Matute et al. did not directly study the potential differences between causal and predictive questions in that study, there are reasons to believe that those two types of questions could produce differential judgments (e.g., Cheng, 1997; Miller & Matute, 1996). For example, a predictive test question (e.g., *To what degree do you expect the outcome to occur in this particular trial?*) seems to refer to a particular time and context and could differ from one time to another (e.g., a light might predict food in this time and context; the same light might predict nothing at a different time or context). However, a causal question (e.g., *To what degree do you think that C is the cause of O*?) seems to imply a more general relationship between the cue and the outcome that does not vary as much with time and context. If this were true, it would be easier to observe extinction with predictive than with causal questions. Finally, a contiguity (or co-occurrence) question (e.g., *To what degree would you say that C was followed by O, even by mere chance?*) should show that regardless of the degree of extinction shown with the other two questions, participants are aware that *C* was followed by *O* in 50% of the trials. In other words, the knowledge that *C* was followed by *O* during the first phase should not be lost during the second phase in which *C* is never followed by *O* even though extinction might be observed as a function of some testing conditions. This was tested in Experiments 1A–1C.

Another variable that might affect the observation of trial-order effects is whether participants are requested to make their judgments as they are learning or only after the training session is finished. Several researchers have found that trial-by-trial responses are more sensitive to recency effects, whereas global responses given at the end of a sequence tend to integrate the information received throughout the session. The influence of this variable has generally been studied in the area of belief revision (e.g., Hastie & Park, 1986; Hastie & Pennington, 1995; Hogarth & Einhorn, 1992), but it has been applied to the area of causal judgments as well (e.g., Catena, Maldonado, & Cándido, 1998; Wasserman et al., 1996). In general, the results of those studies suggest that in our acquisition–extinction design, extinction should be more readily observed if judgments are requested in a trial-by-trial basis than if they are requested only at the end of training. This was tested in Experiments 1A–1C.

Experiments 1A and 1B

In Experiments 1A and 1B, three groups of participants received the identical acquisition–extinction training but differed on whether they received a predictive, causal, or contiguity question. Presumably, extinction (i.e., recency) judgments should be observed in response to the predictive question. However, a judgment integrating the information from the two training phases should be observed in response to the contiguity (co-occurrence) and causal questions. The only difference between Experiments 1A and 1B was that in Experiment 1A judgments were required only at the end of the training session (i.e., global response mode) whereas in Experiment 1B judgments were required on a trial-by-trial basis. As noted in the introduction, in addition to the type of question used at test, the response mode (global vs. trial-by-trial) could also affect the participants' sensitivity to recent trials.

Method

Participants and apparatus. Fifty-seven undergraduate students from Deusto University volunteered for Experiment 1A. Random assignment of participants resulted in 17 participants in Group G.contiguity, 20 participants in Group G.causal, and 20 participants in Group G.predictive.

Forty-seven undergraduate students from Deusto University volunteered for Experiment 1B. Random assignment of participants resulted in 17 participants in Group T.contiguity, 14 participants in Group T.causal, and 15 participants in Group T.predictive (in Experiment 1A, *G* refers to *global;* in Experiment 1B, *T* refers to *trial by trial*).

The experiments were run using personal computers. For each experiment, participants were run simultaneously and seated about 1.5 m apart from each other. Each participant was exposed to a different experimental condition than the two adjacent participants. The experiments were performed with the allergy task (e.g., Wasserman, 1990).

Procedure. Table 1 depicts the design summary of Experiments 1A–1C. The G groups in Table 1 correspond to Experiment 1A, the T groups correspond to Experiment 1B. In each of the two experiments, three groups of participants received identical training consisting of acquisition followed by extinction. Groups differed in that they received different types of questions (i.e., contiguity, causality, or predictiveness). In Experiment 1A, these questions were presented to all three groups once all the information had been presented (i.e., global mode). In Experiment 1B, the questions were presented on a trial-by-trial basis.

In both experiments, the computer showed the records of fictitious patients, one patient per trial. Participants saw two cards per patient (i.e., per trial). The first card indicated that the patient had taken a fictitious medicine, Dugetil. The second card indicated whether the patient had

Table 1
Design Summary of Experiments 1A–1C

Group	Response mode	Phase 1	Phase 2	Test
		Experiment 1A		
G.predictive	G	*C*–*O*	*C*–no*O*	Predictive
G.causal	G	*C*–*O*	*C*–no*O*	Causal
G.contiguity	G	*C*–*O*	*C*–no*O*	Contiguity
		Experiment 1B		
T.predictive	T	*C*–*O*	*C*–no*O*	Predictive
T.causal	T	*C*–*O*	*C*–no*O*	Causal
T.contiguity	T	*C*–*O*	*C*–no*O*	Contiguity

Note. The cue, *C*, is a fictitious medicine; *O*/no*O* indicate the presence or absence of the outcome, an allergic reaction. Twenty trials of each phase were presented. Groups differ on the type of question that they received and on whether the response mode was global (G) or trial by trial (T). Experiment 1C replicated the critical findings in Experiments 1A and 1B using a 2 (response mode: global vs. trial by trial) × 2 (type of question: causal vs. predictive) design.

developed an allergic reaction to the medicine. That is, the first card corresponded to the cue, *C*; the second card corresponded to the outcome, *O*.

In all cases, Phase 1 consisted of 20 trials of *C* followed by *O*, and Phase 2 consisted of 20 trials of *C* followed by no*O*. A screen showing the sentence "Press any key to see the next patient" separated the records for different patients (i.e., trials). A translation from Spanish of the instructions used in these experiments reads as follows:

> *Imagine you are an allergist who wants to study to what degree the consumption of a medicine called Dugetil causes, as a secondary effect, an allergic reaction. The medical records of a series of patients will be presented. Based on them, you will make your predictions. For each patient, you will first see a card that tells you whether that patient has taken Dugetil. 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.*
>
> *At some points during the experiment, you will have to indicate to what degree you think that a particular patient is going to develop the allergic reaction.*

In Experiment 1A, after the 40 training trials were completed, the test trial showed one more card of a patient that had taken Dugetil. This card was identical to the ones used during training, but now, a test question and a rating scale appeared at the bottom of the screen. This test question was the only variable that we manipulated. These questions were as follows (translated from Spanish): *To what degree would you say that the patients that have taken Dugetil have developed, even by mere chance, the allergic reaction?* (contiguity question); *To what degree do you think that Dugetil is the cause of the allergic reaction?* (causal question); and *To what degree do you think that this patient will develop the allergic reaction?* (predictive question). Below the question, the following phrase was displayed: "Please, give a number between 0 and 100, with 0 being absolutely not, and 100 being absolutely." Participants introduced a number by using the *up* and *down arrow* keys and the *Enter* key on the computer keyboard.

Experiment 1B was identical to Experiment 1A except that judgments were required not only in the test trial but also in each of the 40 training trials. The questions and the rating scale used in every trial were identical to those used in the test trial of Experiment 1A.

Results and Discussion

The critical judgments at test are given in Figure 1 for Experiment 1A and in Figure 2 (right panel) for Experiment 1B. For reference, Figure 2 also shows the mean judgments during training in Experiment 1B (left panel).

As can be seen in Figure 1, no differences were observed as a function of test question in Experiment 1A, which used a global response mode. This was confirmed by a one-way analysis of variance (ANOVA), which showed no main effect of type of question in Experiment 1A, $F(2, 54) = 0.91$, $p > .05$. All three groups in this experiment gave an intermediate judgment, suggesting an absence of extinction when the response mode was global.

The results of Experiment 1B, however, showed differences between the different types of questions at the time of testing, and good evidence of extinction was now observed in the group that received the predictive question (see Figure 2). This was confirmed by a one-way ANOVA, which showed a main effect of type of question in Experiment 1B, $F(2, 43) = 4.27$, $p < .05$, and by planned comparisons, which showed that, as expected, extinction was evident in Group T.predictive as compared with Group T.contiguity, $F(1, 43) = 8.53$, $p < .01$. That is, participants in Experiment 1B were able to make the right prediction for the test patient according to the temporal context in which they were being tested (i.e., no one patient was developing the allergic reaction by the end of Phase 2). At the same time, they were also able to integrate the information provided during the two stages when they were asked not to make a prediction for a particular patient but to rate the degree with which *C* and *O* had occurred together during training (i.e., contiguity question).

Figure 2 also shows a greater tendency to integrate the information when the question is causal than when it is predictive. However, the difference between Group T.causal and the other two groups did not reach statistical significance (both *p*s $> .05$). That is, the causal question in this experiment yielded a result that was somehow in between the extinction shown in Group T.predictive and the integrative judgment shown in Group T.contiguity. Thus, the next experiment will try to further clarify this result.

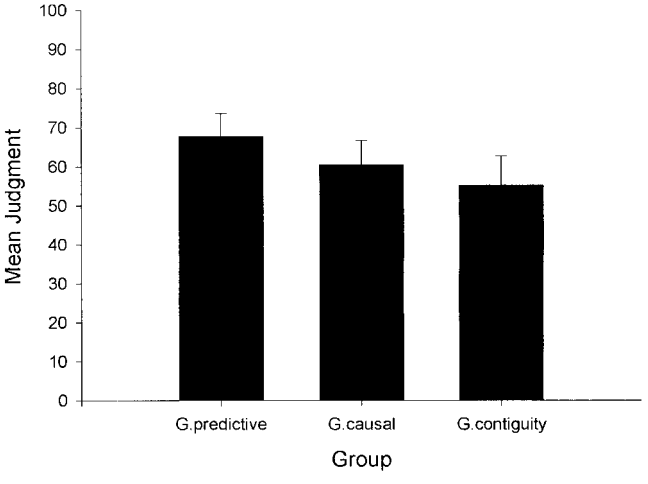

Figure 1. Mean judgment at test for the three groups in Experiment 1A. Error bars represent the standard error of the mean. *G* refers to global response mode.

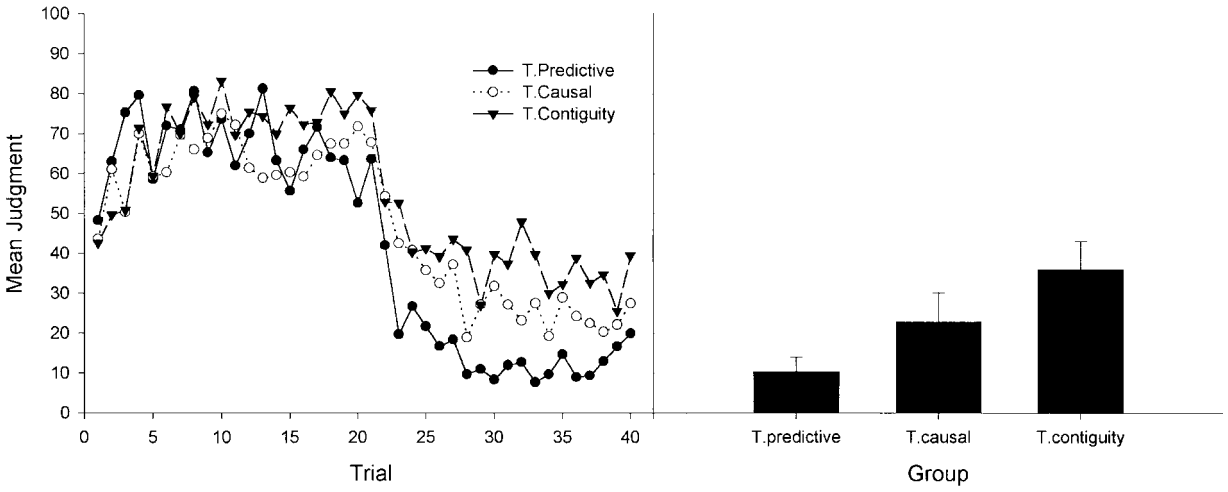

Figure 2. Mean judgment during training (left panel) and testing (right panel) for the three groups in Experiment 1B. Error bars represent the standard error of the mean. *T* refers to trial-by-trial response mode.

Experiment 1C

The results of Experiments 1A and 1B, taken together, suggest that a trial-by-trial response mode and a predictive question are both important to observe the recency effects predicted by associative theories as opposed to the integrative effects predicted by statistical theories. However, because these conclusions are drawn from a comparison between two different experiments, our next step was to assess the influence of those two variables within the same experiment.

Experiment 1C used a 2 (response mode: global vs. trial by trial) × 2 (type of question: predictive vs. causal) design. We did not include the contiguity question in this experiment because the results of Experiments 1A and 1B consistently showed that, regardless of whether the response mode is global or trial by trial, participants are able to give an accurate judgment of the degree to which the two events have occurred together during training.

Although the results of Experiment 1B replicated those of Experiment 1A with respect to the contiguity question and extended them with respect to the predictive question, they were not clear cut with respect to the causal question. This question showed a tendency toward a more integrative judgment than that of the predictive question, but it was not clear how reliable this tendency was. It might be worth noting that in Experiments 1A and 1B the causal question was presented, just as the other questions, in the same screen in which the cue (i.e., *This patient has taken Dugetil*) was presented. Although this makes sense with respect to the predictive question (i.e., *To what degree do you think that this patient will develop the allergic reaction?*), the causal question was not intended to refer to one particular patient but to the causal relationship between Dugetil and the allergic reaction in general. By asking this question in the same screen in which the information that a particular patient had taken Dugetil was presented, we might have confused participants: This screen might have suggested to some of them that we were asking whether Dugetil would be followed by the allergic reaction in this particular patient. If this were the case, the causal question might have been interpreted as predictive by some participants and as causal by other participants. Therefore, to avoid this possible misinterpretation and to make it more clear to participants what each question meant, the causal group in Experiment 1C will first see the cue card, then the outcome card, and only after they have seen the two cards will they be asked about the causal relationship between the cue and the outcome.

Method

Participants and apparatus. Sixty-eight undergraduate students from Deusto University volunteered for the study. None of the participants had taken part in any related experiment. Random assignment of participants resulted in 16 participants in Group T.predictive, 18 participants in Group T.causal, 17 participants in Group G.predictive, and 17 participants in Group G.causal. The participants were run in individual cubicles.

Procedure and design. The experiment was run again with the allergy task, and the design was a combination of Experiments 1A and 1B. The causal and predictive groups in Table 1 are the ones that were replicated in this experiment. Participants received again 20 *C*–*O* trials followed by 20 *C*–no*O* trials, as in the previous experiments. However, there were two groups trained in the trial-by-trial response mode (T groups) and two other groups trained in the global response mode (G groups). In both response-mode conditions, half of the participants received predictive questions, the other half received causal questions.

The main procedural difference with respect to Experiments 1A and 1B concerns the causal groups. In these groups, participants were now requested to introduce their judgment only after they had seen both the cue and the outcome screen, so that they could not assume that they were supposed to make a predictive judgment for a particular patient. In the same vein, the second paragraph of the instructions that participants received at the beginning of the experiment was also modified as follows for the causal groups: *At some points during the experiment, you will have to indicate to what degree you think that Dugetil is the cause of the allergic reaction.*

Our presenting the predictive question before the outcome and the causal question after the outcome implies that participants in the causal groups would receive one more training trial than the predictive groups before making their final test judgment. For this reason, the causal groups in this experiment were tested in the 40th trial rather than in the 41st trial.

Results and Discussion

The left panel of Figure 3 shows the training data for the two groups that responded during training (i.e., the T groups). The right panel shows the critical data, that is, the mean judgment at test for all groups.

The results of this experiment replicated the findings of Experiments 1A and 1B that extinction is most clearly observed at test when a predictive question and a trial-by-trial response mode is used, and that, by contrast, global and causal questions tend to favor more integrative judgments. A 2 (response mode) $\times$ 2 (type of question) ANOVA on the final judgments during testing yielded a main effect of type of question, $F(1, 64) = 8.23$, $p < .01$, a main effect of response mode, $F(1, 64) = 14.79$, $p < .01$, and no interaction, $F(1, 64) = 0.74$, $p > .05$. Planned comparisons between the two G groups replicated the results of Experiment 1A, in that no differences were observed between predictive and causal judgments when the response mode was global, $F(1, 64) = 2.01$, $p > .05$. In that case, participants tended to integrate the information received during the two training phases regardless of whether they were asked a predictive or a causal question. However, planned comparisons between the two T groups replicated and clarified the results of Experiment 1B, in that they show that the predictive question produced lower judgments than the causal question when the response mode was trial by trial, $F(1, 64) = 6.95$, $p < .05$. Thus, questions that are trial by trial and predictive seem to favor the observation of recency effects, whereas a global response mode or a causal question seem to favor judgments that integrate the two training phases.

Experiment 2

The results of Experiments 1A–1C, taken together, suggest that participants making causal or global judgments tend to integrate the information received through the training session, whereas those making predictive judgments on a trial-by-trial basis tend to be more sensitive to the information received during recent trials. Thus, both the response mode and the question type seem to be critical factors that modulate the observation of one or the other type of judgment.

Of importance, it seems that recency effects, which are often used to discriminate between statistical and associative predictions (see, e.g., Shanks et al., 1996), tend to occur only in predictive trial-by-trial situations. Before drawing such a general conclusion, however, it is important to make sure that our results are not restricted to the acquisition–extinction design that we have been using in Experiments 1A–1C. For this reason, in Experiment 2 half of the participants received the two training phases in the same order as in the previous experiments, whereas the other half received the two training phases in the reverse order (i.e., the *C*–no*O* trials presented before the *C*–*O* trials). Orthogonally, the response mode was either global or trial by trial for half of the participants within each group. The question used was always predictive. If the above conclusions are correct, trial order should affect predictive judgments in the T groups (with the group receiving the *C*–*O* trials during the second phase giving a higher judgment at test than the group receiving *C*–no*O* trials during the second phase), but not in the G groups.

Method

Participants and apparatus. Eighty undergraduate students from Deusto University volunteered for the study. None of the participants had taken part in any related experiment. Sixty-five participants were run in a large computer room, as in Experiment 1A. The remaining participants were run in individual cubicles, as in Experiment 1C. In both testing conditions, the participants were randomly distributed across the four experimental groups. This resulted in 20 participants per group.

Procedure and design. Table 2 depicts the design summary of this experiment. Two groups received trial-by-trial training; the two other groups received global training. Half of the participants in each condition received the same acquisition–extinction training of the previous experiments, that is, *C* was followed by *O* in all trials during Phase 1 and never during Phase 2 (Groups G.1.0 and T.1.0). For the other half of participants, the order of phases was reversed, that is, *C*–no*O* trials were followed by

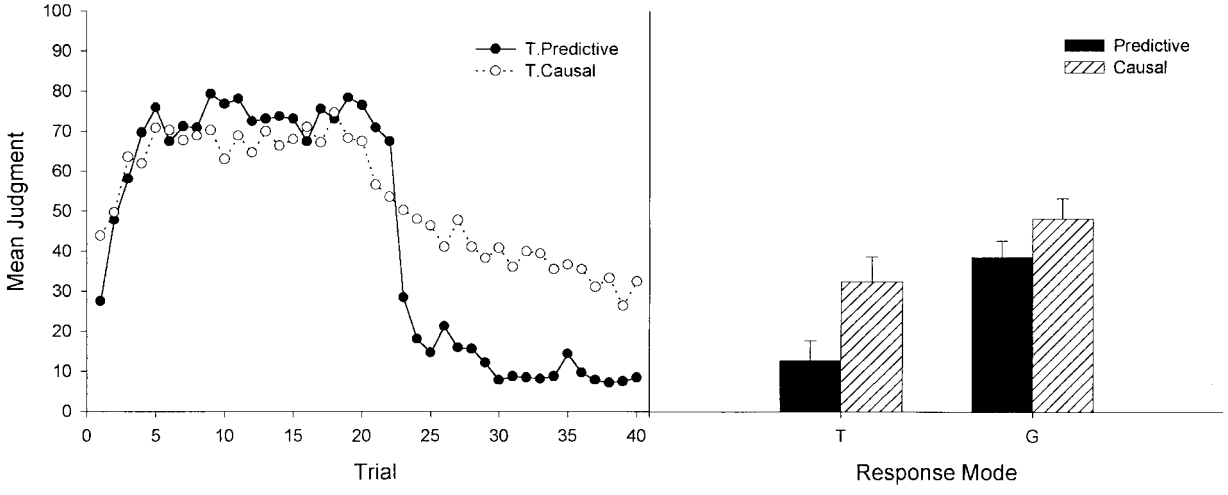

Figure 3. The left panel shows the mean judgment during training for the two trial-by-trial groups in Experiment 1C. The right panel shows the mean judgment at test for all four groups as a function of response mode (G vs. T) and type of question (predictive vs. causal). Error bars represent the standard error of the mean. *G* refers to global response mode; *T* refers to trial-by-trial response mode.

Table 2
Design Summary of Experiment 2

Group	Response mode	Phase 1	Phase 2	Test
G.1.0	G	*C*–*O*	*C*–no*O*	Predictive
G.0.1	G	*C*–no*O*	*C*–*O*	Predictive
T.1.0	T	*C*–*O*	*C*–no*O*	Predictive
T.0.1	T	*C*–no*O*	*C*–*O*	Predictive

Note. The cue, *C*, is a fictitious medicine; *O*/no*O* indicate the presence or absence of the outcome, an allergic reaction. Groups differ in their response mode (G vs. T), as well as in the order of the two training phases (e.g., 1.0 means that the outcome always occurred during Phase 1 and never during Phase 2; 0.1 means the reverse order of trials). G = global; T = trial by trial.

C–*O* trials (Groups G.0.1 and T.0.1). In all cases, 20 trials of each type were presented, and the type of question was always predictive.

Results and Discussion

The left panel of Figure 4 shows the training data for the two groups that responded during training (i.e., the T groups). The right panel shows the critical data, that is, the mean judgment at test for all groups. As expected, trial order affected predictive judgments at test in the two trial-by-trial groups but not in the global groups. These impressions were confirmed by a 2 × 2 ANOVA, which showed a main effect of trial order, $F(1, 76) = 20.21$, $p < .01$, no main effect of response mode, $F(1, 76) = 0.02$, $p > .05$, and a Trial Order × Response Mode interaction, $F(1, 76) = 10.11$, $p < .01$. As shown in Figure 4, the results at test depended strongly on trial order when the response mode was trial by trial but not when it was global. Planned comparisons showed a significant difference between the two T groups, $F(1, 76) = 29.45$, $p < .01$, but no significant difference between the two G groups was observed, $F(1, 76) = 0.87$, $p > .05$. Thus, as expected, the two G groups integrated the information received during the two training phases, regardless of trial order, whereas the two T groups tended to respond during testing according to the information received during the latest trials. This suggests that the results of Experiments 1A–1C are not restricted to the acquisition–extinction design that we used, and can be generalized to other situations in which the order of trials is not randomly distributed through the training session.

Experiment 3

As expected, the differences that we have observed among the several conditions tested in Experiments 1A–1C and 2 suggest that participants can respond differently as a function of how they interpret the test question. Both the trial-by-trial response mode and the predictive wording seem to be interpreted as requiring a specific judgment for a particular case in a particular time and context; however, the global response mode and the causal and contiguity questions seem to be interpreted as demanding a response that integrates the two training phases.

However, there is also a possibility that the differences observed between the global and the trial-by-trial conditions are merely due to a lack of attention in the global conditions. These participants have nothing to do during training but reading the information on the screen and pressing the *Enter* key to go on to the next trial. Thus, it is possible that they simply keep pressing the *Enter* key without even reading the information presented in each trial. The present experiment tested this possibility.

Method

Participants and apparatus. Forty undergraduate students from Deusto University volunteered for the study. None of the participants had taken part in any related experiment. The participants were run in a large computer room, as in Experiment 1A, and were randomly distributed across the four experimental groups. This resulted in 11 participants in Group G, 11 participants in Group T, 9 participants in Group G.predict, and 9 participants in Group G.read.

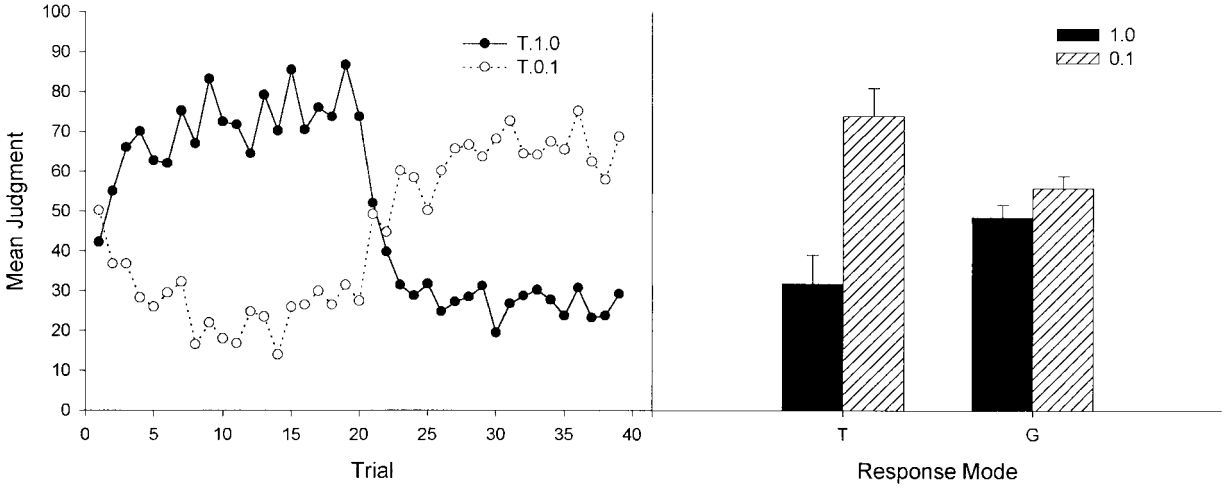

Figure 4. The left panel shows the mean judgment during training for the two trial-by-trial groups in Experiment 2. The right panel shows the mean judgment at test for all four groups as a function of response mode (G vs. T) and trial order (1.0 vs. 0.1). Error bars represent the standard error of the mean. *G* refers to global response mode; *T* refers to trial-by-trial response mode.

Procedure and design. Groups G and T replicated the global and the trial-by-trial conditions of Experiments 1A–1C, with the type of question being predictive in all cases. To test whether a lack of attention could be responsible for the results of the global conditions, two other global groups were used in this experiment that emitted a judgment only at test but that had to type a yes/no response in each training trial. Group G.predict was required to give a yes/no prediction for the outcome when the cue was presented in each trial. The wording of this question was *Will this patient develop the allergic reaction? (y/n).* Group G.read was simply forced to read the screen: While the information on the allergic reaction was still visible, these participants were required to give a yes/no response to the question *Did this patient develop the allergic reaction? (y/n).* In this group, the program did not go on to the next trial until the information was read and the correct yes/no response was typed.

Results and Discussion

The left panel of Figure 5 shows the training data from the group that emitted judgments during training (Group T). The insert in this figure shows the percentage of "yes" predictions during training for Group G.predict. This insert shows that, like participants in Group T, participants in Group G.predict were also paying attention to the information presented in the screen: they predicted the occurrence of the outcome during Phase 1 and its absence during Phase 2. Finally, the percentage of "yes" responses in Group G.read is not shown in the figure, but these participants were also paying attention to the screen and their responses were 100% correct (recall that in this group the program did not continue to the next trial until the correct response had been typed).

The right panel of Figure 5 shows the critical data in this experiment, that is, the mean judgment at test for all groups. As can be seen in this panel, the three global groups behaved homogeneously and differed from the trial-by-trial group. This was confirmed by a one-way ANOVA, which showed a main effect of group, $F(3, 36) = 10.76$, $p < .01$, and by planned comparisons, which showed that, as expected, extinction was evident in Group T as compared with Group G, $F(1, 36) = 18.96$, $p < .01$, Group G.predict, $F(1, 36) = 20.81$, $p < .01$, and Group G.read, $F(1, 36) = 22.51$, $p < .01$. This shows that the results of the previous experiments cannot be attributed to a lack of attention in the global conditions.

Experiment 4

Experiments 1A–3 have shown that both the type of question and the response mode affect the way in which participants make their judgments at the end of training. With regard to the type of question, it is not that surprising that participants receiving a predictive question tend to rely on recent information to make their prediction because this is probably an efficient way to make accurate predictions in a changing environment. The large amount of data existent on extinction in animal research can also be regarded as a predictive trial-by-trial situation in which, as in the present experiments, participants tend to make their predictions on the basis of the most recent information received. At the same time, it is not that strange that participants that were asked to make a causal rather than a predictive judgment tended to consider a greater amount of information that included also the early trials. As suggested by the results to the contiguity questions in Experiments 1A and 1B, participants were aware of the probability of the outcome occurring in the presence of *C* being .5 (i.e., 1 during the early trials, 0 during the latest trials). However, they also were able to predict a probability of 0, or close to 0, for the next trial when the question was predictive, or to infer a causal relation of an intermediate level between the two events when they were asked to make causal judgments. That is, participants seemed to be able to use the acquired information in a flexible way as a function of the test question that they received (see also Matute et al., 1996, for a similar finding).

More puzzling is the response-mode effect. Why should participants respond differently as a function of whether they are responding in every trial or at the end of training? The idea that

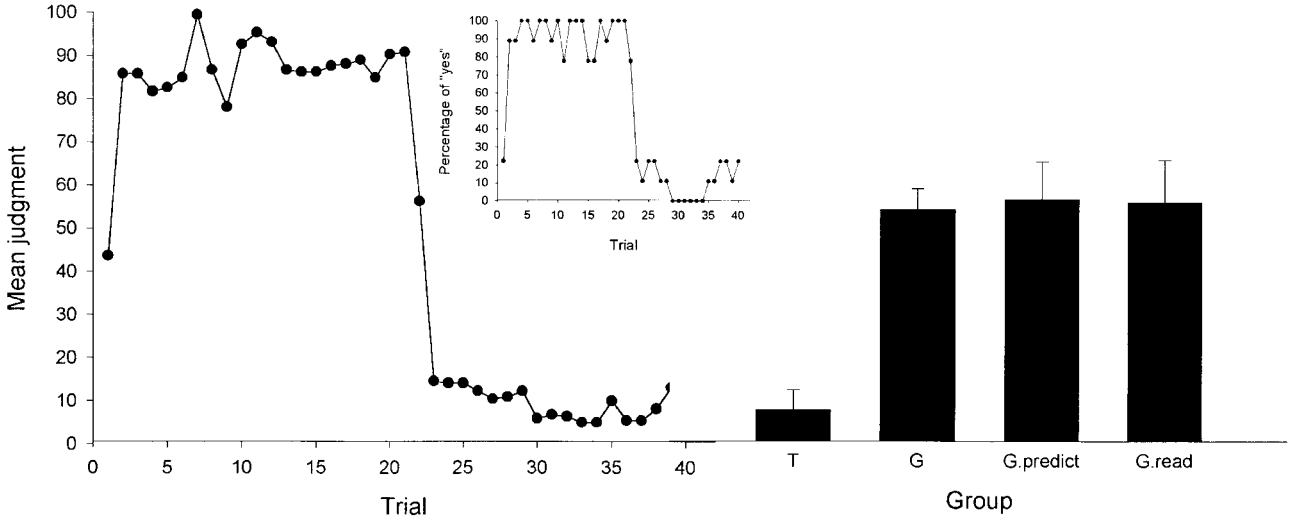

Figure 5. The left panel shows the mean judgment during training for Group T in Experiment 3. The insert shows the percentage of "yes" predictions during training for Group G.predict. The right panel shows the mean judgment at test for all four groups. Error bars represent the standard error of the mean. *G* refers to global response mode; *T* refers to trial-by-trial response mode.

participants giving global judgments store all trials whereas those giving trial-by-trial judgments do not (Catena et al., 1998; Hastie & Park, 1986; Hastie & Pennington, 1995; Hogarth & Einhorn, 1992) is an appealing one. However, this cannot explain the difference we observed among trial-by-trial participants that received different test questions: if they had stored only the recent trials, then they should have given judgments close to 0 in all cases (except for Group T.0.1 in Experiment 2).

Our observation that participants receiving contiguity or causal questions tended to integrate the information of the two phases at test even when their response mode was trial by trial, suggests that the locus of their integration was at the response stage rather than at the acquisition stage. That is, like the global participants, trial-by-trial participants must also have had all the information at the end of training to respond at test one way or another as a function of how they interpreted the test question. If this is true, the information could probably be integrated at the end of training, not only as a function of the response mode and the type of question, but also as a function of many other testing demands. Indeed, anything that during testing would suggest to participants that an integrative judgment would be more appropriate than the default (i.e., recency) judgment in a trial-by-trial situation should produce an integrative response. This view was tested in the present experiment.

In this experiment, three groups of participants received the conditions that should produce the strongest recency effect, that is, predictive questions in a trial-by-trial response mode. Our primary manipulation consisted of inserting an instructional screen just before testing. The purpose of this screen was to suggest that the test trial should not be interpreted as an additional Phase 2 trial (for which a recency response would be appropriate), but as a whole new phase. In this way, all they had learned during the two training phases, rather than only the information from Phase 2, should become relevant at test. If the effects we have been observing in the previous experiments are response effects, participants trained with the trial-by-trial mode and predictive questions should be able to produce integrative judgments if they interpret the test phase in this way. By contrast, if these effects are due to the trial-by-trial participants failing to store the information received through the training session, these participants will be unable to give an integrative judgment even if the test instructions suggest that it will be appropriate to do so.

Method

Participants and apparatus. Sixty-eight undergraduate students from Deusto University volunteered for the study. None of the participants had taken part in any related experiment. They were randomly assigned to the various conditions, which resulted in 17 participants per group. The participants were run in individual cubicles, as in Experiment 1C.

Procedure and design. Table 3 shows the design summary for this experiment. Three groups were exposed to a trial-by-trial response mode and one group was exposed to a global response mode. Groups G.20 and T.20 replicated the global and trial-by-trial conditions of the previous experiments, with the question being predictive in all cases. That is, they received 20 *C–O* trials followed by 20 *C*–no*O* trials. Two other groups (Groups T.20.i and T.33.i) were exposed to a trial-by-trial mode, but at the end of the training sequence, just before testing, they received the following instruction (translated from Spanish):

> *You now should have enough information about the relationship between Dugetil and the allergic reaction. Now imagine that a patient of yours has taken Dugetil and she asks you if she is going to develop the allergic reaction. You will have to tell this person to what degree you think she is going to develop the allergic reaction so that she can decide whether or not she should give up the treatment with Dugetil.*

One of the instructed groups (Group T.20.i) received the same 20 *C–O* and 20 *C*–no*O* trials as the T.20 group. If the instruction that they received before testing was enough for them to integrate what they had learned during the two training phases, they should give a judgment close to 50 in our 0–100 scale, that is, significantly higher than that of Group T.20 and similar to that of Group G.20. However, there was a possibility that participants in Group T.20.i gave a judgment of 50 not because they integrated the information of the two phases, but simply because they wanted to say, "I don't know; so fifty–fifty." To avoid this possible interpretation of the expected judgment of 50 in this group, the other instructed group (Group T.33.i) received 33, rather than 20, extinction trials. Thus, if the instructed groups were integrating the information that they had received and were giving an accurate judgment rather than a mere "I do not know" response, the judgments of these two instructed groups should differ from each other.

Results and Discussion

The left panel of Figure 6 shows the training data for the three groups that responded during training (i.e., the T groups). The critical results of this experiment are depicted in the right panel for all four groups. As expected, Groups T.20 and G.20 replicated during testing the basic response-mode effect, but more important,

Table 3
Design Summary of Experiment 4

Group	Response mode	Phase 1	Phase 2	Instruction	Test
G.20	G	20 *C–O*	20 *C*–no*O*	No	Predictive
T.20	T	20 *C–O*	20 *C*–no*O*	No	Predictive
T.20.i	T	20 *C–O*	20 *C*–no*O*	Yes	Predictive
T.33.i	T	7 *C–O*	33 *C*–no*O*	Yes	Predictive

Note. The cue, *C*, is a fictitious medicine; *O*/no*O* indicate the presence or absence of the outcome, an allergic reaction. Groups G.20 and T.20 are intended to replicate the basic response-mode effect of Experiments 1–3. Group T.20.i was included to assess the prediction that an instruction received just before testing would be sufficient to produce an integrative judgment in participants that had received trial-by-trial training with predictive questions. Group T.33.i was a control group that was identical to Group T.20.i except that it received a larger number of extinction trials (i.e., 33 rather than 20). G = global; T = trial by trial.

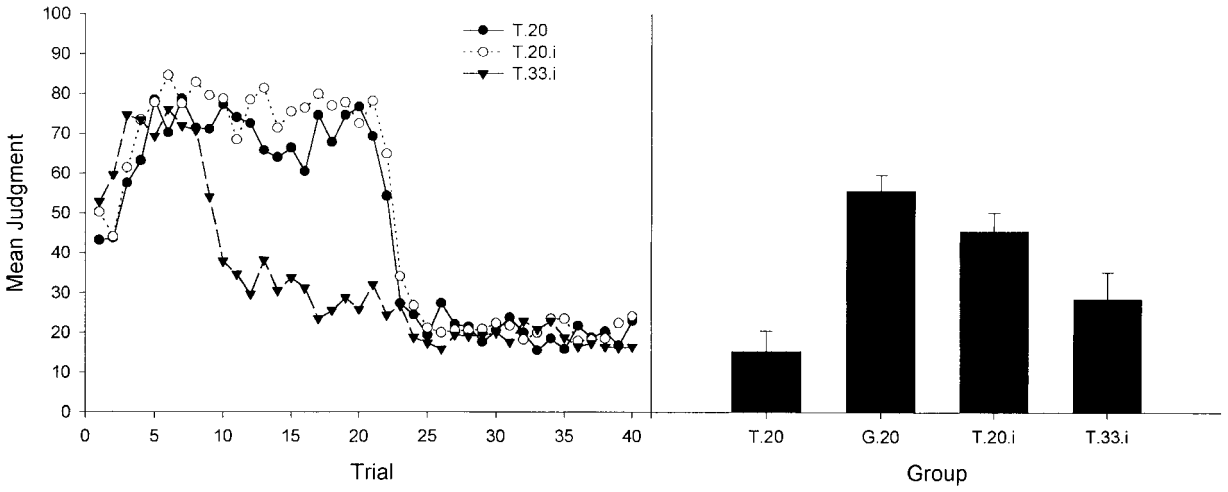

Figure 6. The left panel shows the mean judgment during training for the three trial-by-trial groups in Experiment 4. The right panel shows the mean judgment at test for all four groups. Error bars represent the standard error of the mean. *G* refers to global response mode; *T* refers to trial-by-trial response mode.

Group T.20.i behaved as the global group rather than as a trial-by-trial group. Participants in Group T.20.i were able to integrate the information from the two training phases when giving their final judgment. Moreover, the judgment of Group T.20.i cannot be interpreted as an "I don't know" response, because the judgment of this group differed from that of Group T.33.i, which also showed an integrative, but lower, judgment.

These impressions were confirmed by a one-way ANOVA on the mean judgments at test, which showed a main effect of group, $F(3, 64) = 11.94$, $p < .01$. Planned comparison between Groups T.20 and G.20 showed a basic response-mode effect, with judgments in Group T.20 being significantly lower than those in Group G.20, $F(1, 64) = 30.36$, $p < .01$. Moreover, when an instruction to integrate the information was inserted in the T.20 group before testing (i.e., Group T.20.i), the judgment of the T.20 group became significantly increased in Group T.20.i, $F(1, 64) = 16.91$, $p < .01$. Indeed, the judgment of Group T.20.i is almost as high as that reached by Group G.20, $F(1, 64) = 1.95$, $p > .05$. This suggests that a global response mode is not necessary for participants to integrate what they had learned in the two phases. Finally, the difference between Group T.20.i and Group T.33.i, $F(1, 64) = 5.36$, $p < .05$, indicates that these two groups are accurately integrating the information rather than giving an "I don't know" response.

Thus, the response of Group T.20.i replicates the results of the global groups in the previous experiments. In those cases, participants did not show recency effects. Instead, they regarded the two training phases as relevant when making their final judgment. The same result was now observed in Group T.20.i through a different manipulation. Therefore, it seems that it is not the frequency with which judgments are requested, but how the demands of the test phase are interpreted, that produces one or the other result.

General Discussion

The experiments presented here show that there are some testing conditions that favor the recency effects predicted by associative theories, whereas there are other testing conditions that favor the absence of trial-order effects predicted by statistical theories. There are probably many other conditions, in addition to the ones we have tested here, that could also favor the observation of one or the other type of judgment. Nevertheless, summarizing our results, these are the variables that we have found to be critical:

First, the type of question used to assess participants' judgments was a critical variable. Among the different types of questions that we assessed, the contiguity question showed that participants accurately store all the information and can give accurate judgments on the degree with which *C* and *O* occurred together through the session, regardless of whether the response mode is global or trial by trial (Experiments 1A and 1B). However, participants tended to use only the most recent information when the test question was predictive, particularly when the response mode was trial by trial (Experiments 1B–4). The causal questions, however, tended to yield more integrative judgments than the predictive questions. Causal judgments do not normally refer to a particular case in a particular situation and time, as predictive questions do, but to a more general relationship between the cue and the outcome (Experiments 1B and 1C). Thus, in general, extinction and trial-order effects tend to be observed more readily in predictive situations.

Second, we observed a very strong response-mode effect that was evidenced in the trial-by-trial mode being generally more sensitive to recent information than the global mode, particularly when predictive questions were used (Experiments 1A–4). Moreover, Experiment 3 showed that the response-mode effect was not due to a lack of attention in the global conditions, and Experiment 2 showed that this effect is not restricted to the acquisition–extinction design that we used in all the other experiments: In Experiment 2 we tested the reverse order of trials, and the results were consistent with the results of the other experiments. That is, trial order affected participants' judgments in the trial-by-trial response mode but not in the global response mode.

Third, informing participants just before testing that the test trial should not be interpreted as an additional Phase 2 trial was sufficient for the trial-by-trial participants that were receiving predictive questions to make an integrative judgment rather than their

default recency-based judgment (Experiment 4). Thus, as robust as the recency effect might be in the trial-by-trial conditions, it seems that it can be counteracted not only by the test question (Experiments 1B and 1C) but also by simply separating the test trial from the Phase 2 trials, therefore making all previous phases equally relevant at the time of testing (Experiment 4). This suggests that the response-mode effect has less to do with the frequency with which judgments are requested than with how participants interpret the demands of the environment at the time of testing.

At first glance, our results might seem inconsistent with previous results reported by Wasserman et al. (1996, Experiments 2 and 3). In some conditions, their participants received a positive contingency during the first half of the study and a negative contingency during the second half. In some other conditions, the order of trials was reversed. Moreover, some participants responded on a trial-by-trial basis, whereas some responded on a global mode. Wasserman et al. did not report trial-order effects. However, the purpose of their experiments was different from ours, and they used an overall ΔP of 0 and a rating scale that went from -10 to 10. Thus, all judgments were close to 0 at asymptote, and the differences we observed could not be observed in their study (note that if the overall ΔP is 0, there is no reason to expect our results: the predictions of associative and statistical models are coincident at asymptote in that case). However, their training data showed sensitivity to the order of trials, and the trial-by-trial pattern that they reported was very similar to the one observed in the present series of studies: participants having the positive contingency first gave higher judgments during the first half of the study, and this was reversed for participants having the negative contingency first. Translated to their design and scale, our results would probably be very similar to theirs.

Potential Explanations

Trial-order effects are frequently regarded as one of those phenomena that can best be used to discriminate between the statistical and associative predictions (e.g., Allan, 1993; López et al., 1998): Statistical theories predict response effects that should result in integrative judgments, whereas associative theories, such as that of Rescorla and Wagner (1972), predict acquisition effects that should result in recency effects (i.e., extinction). However, because we observed both trial-order effects and an absence of trial-order effects as a function of several testing variables, none of those two sets of theories can directly provide a unified account of the observed results.

One possibility would be to consider the belief revision models. These models predict that information is processed and stored differently as a function of the response mode, with only participants exposed to the global mode storing all the information (Catena et al., 1998; Hogarth & Einhorn, 1992; Pennington & Hastie, 1992). These models could at first glance explain the response-mode effect that we observed in Experiments 1–3. However, as shown above, our results suggest that the response-mode effect is not due to differential processing or storing of the information but to differential responding at test. In Experiment 4, if participants in Groups T.20.i and T.33.i had not stored the information received in all trials, they would not have been able to give an integrative judgment at test. However, an instructional screen presented just before testing was sufficient for these participants to be able to integrate the information. This suggests that the information storing strategies of the trial-by-trial response mode are not different from those of the global response mode.

Indeed, our results show that participants respond one or the other way as a function of test demands, which suggests that any model must allow for flexible use of information once it has been acquired: Sometimes it might be beneficial to make a prediction based on recent information, sometimes it might be more adaptive to integrate the information received at different stages, sometimes it might be beneficial to rely on information received during the early trials. But, to be able to use either the most recent or the oldest information, or all of it, and to respond appropriately to each of the changing environmental demands, it also seems necessary to store as much information as possible (e.g., *C*–*O* and *C*–no*O* information in the present case) in a simple and efficient way.

Perhaps the most obvious possibility would be to extend statistical theories because they at least do store all the information. One of the most popular current statistical models of causal learning is Cheng's (1997) Power PC model. However, Cheng has explicitly argued that her model does not apply to predictive judgments, and those are the only ones that we used in our Experiments 2–4. Nevertheless, Cheng's model was a revision of an older model by Cheng and Novick (1992), which applied to predictive judgments. This older model introduced the notion of focal sets: not all the information is used when computing ΔP, but only the information in the relevant focal set. These focal sets are formed when there is more than one potential predictor of the outcome and the two or more predictors sometimes occur together. However, when there is only one cue, as in the present experiments, the predictions of the focal-set ΔP rule are identical to those of the original ΔP rule. Thus, in the present experiments, the predictions of the focal-set ΔP rule and those of the original ΔP rule are coincident. They predict that participants will use all the *C*–*O* and the *C*–no*O* trials when computing their judgments at test. This allows these theories to explain the integrative judgments observed in some conditions in the present studies but prevents them to explain the trial-order effects observed in the other conditions. Indeed, the lack of an acquisition function prevents these theories to account for any acquisition-related effects, such as trial order, learning curves, and preasymptotic judgments (see also Allan, 1993; Baker et al., 1993; López et al., 1998; Shanks, 1993).

It should be possible, however, to develop statistical models provided with an acquisition function as well as with time-sensitive focal sets. This would probably require the addition of a recency parameter to each focal set, so that at test, participants could decide whether the most recent focal set, the oldest one, or all of them should be taken into account when calculating their judgments. This would also require one to specify how (and when) such time-sensitive focal sets should be created. As others have noted (e.g., Shanks, 1993), one of the main problems in trying to apply the notion of focal sets to a particular experiment is how to define in advance which focal sets will be created in each particular situation.

The alternative possibility is to consider associative theories. As previously noted, the main difference between statistical and associative models is that statistical models focus on the response process and associative models generally focus on the trial-by-trial updating of the associations during acquisition. This updating process allows them to explain the trial-by-trial results. At the same time, however, this updating process prevents them from accounting for the integrative judgments observed in the present

experiments because it results in the destruction of the *C–O* association by the end of the extinction phase (e.g., Rescorla & Wagner, 1972). This problem has been called "catastrophic forgetting" by some researchers (e.g., McCloskey & Cohen, 1989; Ratcliff, 1990). However, there are some exceptions to this rule.

There are some models of animal learning that, unlike the most prevalent approach (e.g., Rescorla & Wagner, 1972), regard extinction and other interference effects as retrieval effects rather than as acquisition effects (e.g., Bouton, 1993). In these cases, extinction does not imply the destruction of the *C–O* association acquired during Phase 1. Instead, the organism acquires two different associations: a *C–O* association during Phase 1, and a *C*–no*O* association during Phase 2. The problem here is that the meaning of the cue becomes ambiguous, and when it is presented during testing, *C* could predict either *O* or no*O*. Thus, it is necessary to specify when extinction will be observed and when it will not. According to Bouton (1993), the (physical or temporal) context in which the test phase takes place plays the disambiguating role. If the test phase is conducted immediately after Phase 2 and in the same physical context, as is usually the case, the most recently acquired association (i.e., *C*–no*O*) is the one that will be activated by the test context. Thus, extinction will be the default result. However, if testing occurs in a novel (temporal or physical) context, the most recently acquired *C*–no*O* association does no longer prevail and good responding will be observed. The many experiments showing spontaneous recovery when a time interval elapses between extinction and testing and renewal of responding when testing occurs in a novel context give support to this view (see e.g., Bouton, 1993, for a comprehensive review).

Bouton's (1993) theory was developed in the area of animal conditioning and cannot directly account for the results observed in the present experiments. However, it could be readily extended to human learning if we assume that in humans the disambiguating role does not need to be restricted to the (temporal or physical) context in which testing takes place. In principle, not only contexts, but also verbal stimuli and other test demands such as response mode or mere instructions can help participants disambiguate whether the *C–O* or the *C*–no*O* association, or the two of them, are relevant at a particular time and context. Apparently, instructions and test questions can also activate one or the other association just as contexts do. Moreover, according to Bouton (1993), spontaneous recovery and renewal of responding occur because the *C*–no*O* association acquired in a particular time and context does not transfer to novel contexts as readily as the *C–O* association does. However, it is also possible that because the novel context cannot be of great help in disambiguating whether *O* or no*O* will occur after *C*, the two associations might become partially activated in such cases by the test context. Thus, each of them could interfere with the expression of the other one. In this case, the combined excitatory and inhibitory strength of these two associations could be responsible for the intermediate degree of responding that is normally observed when testing occurs in a novel (temporal or physical) context. This could also explain the integrative judgments observed in the present research.

For example, the response-mode effect can be understood if we assume that in the T groups the test trial must have been perceived as an additional Phase 2 trial (there was nothing that allowed participants to assume that it belonged to a different phase) and in consequence, the default Phase 2 association (i.e., *C*–no*O*) prevailed at test. The G groups, however, could not perceive the test trial as an additional Phase 2 trial: It was clear that it was a whole new (i.e., test) situation in which they were required to introduce a judgment for the first time after the two training phases had been completed. Therefore, the two associations could become equivalently activated in such testing conditions, thus resulting in an intermediate (or integrative) judgment.

Similarly, the causal and contiguity questions that we used were explicitly worded to suggest that we were not referring exclusively to the latest trials, and in consequence, participants integrated the information from the two training phases (i.e., the two associations become activated). However, the predictive test question explicitly referred to a particular patient in a particular time and context (i.e., at the end of Phase 2). Thus, unless responding on a global mode (Experiments 1–4) or receiving instructions to the contrary (Experiment 4), participants receiving the predictive question tended to respond as if the test trial belonged to Phase 2 (i.e., "if patients are no longer developing the allergic reaction and nothing has changed for the last 20 trials, why should I predict that the next patient should be different?") The observed type-of-question effect provides a replication of the basic finding reported by Matute et al. (1996) in that associations seem to be acquired on the basis of mere contiguity, and these associations can then be flexibly used at test. As a function of how the test phase is interpreted, one or more associations can become activated. In consequence, interference can be observed when more than one association becomes activated at test.

In the same vein, the instructional manipulation that we used in Experiment 4 could readily be interpreted as serving the purpose of separating the test trial from the Phase 2 trials. That is, these instructions told participants in the T groups that the next trial (i.e., the test trial) was not to be regarded as an additional extinction trial, but as a new type of trial. In consequence, both the *C–O* and the *C*–no*O* associations become similarly activated and interfered with the expression of the other one.

Conditions That We Have Not Tested

There are many conditions that we have not tested. For example, we have not tested how these variables apply to more complex manipulations, such as those in which cues are presented in compound (as in forward and backward blocking), or those in which other trial types are used in addition to the ones we have used here (e.g., no*C–O* and no*C*–no*O* trials, instead, or in addition to, the *C–O* and *C*–no*O* trials that we used). These are interesting questions for future research because, for example, the observation of both forward and backward blocking (i.e., evidence for an absence of trial-order effects) in some studies (e.g., Shanks, 1985), along with the observation that backward blocking is weaker and harder to observe (i.e., evidence for a trial-order effect) in other studies (e.g., Chapman, 1991), suggest that the conditions that yield to one or the other type of result are not yet clear. It is possible that some of the variables that we have tested here could also have an effect on those compound conditions as well. Moreover, the results of Catena et al. (1998), show that the response mode affects the observation of trial-order effects in situations in which the four trial types (*C–O*, *C*–no*O*, no*C–O*, and no*C*–no*O*) are manipulated. Although they did not test the effect of the type of question in their experiments, it is possible that both the type of question and the instructions presented just before testing could also affect condi-

tions in which the four trial types, rather than the two used here, are distributed throughout the experiment.

Finally, other response-mode manipulations, such as asking participants to emit one judgment at the end of each phase, rather than on every single trial, have been used in some studies of human extinction (e.g., Vila, 2000). The results observed under those conditions show sensitivity to the extinction trials. Thus, it seems that as long as there is at least one judgment per phase, participants tend to respond accordingly to the information provided during that phase rather than to respond in an integrative manner. This, again, suggests that trial-order effects depend on whether the test question is interpreted as referring to the most recent phase or to the whole training session.

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

Baker, A. G., Mercier, P., Vallée-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 414–432.

Bouton, M. E. (1993). Context time and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80–99.

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 effects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 837–854.

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

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 judgment required modulate cue competition? *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 53*(B), 193–207.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93,* 258–268.

Hastie, R., & Pennington, N. (1995). Cognitive approaches to judgment and decision making. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 32, pp. 1–31). San Diego, CA: Academic Press.

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

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

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.

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.

Miller, R. R., & Escobar, M. (2001). Contrasting acquisition-focused and performance-focused models of acquired behavior. *Current Directions in Psychological Science, 10,* 141–145.

Miller, R. R., & Matute, H. (1996). Animal analogues of causal judgment. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Causal learning* (Vol. 34, pp. 133–166). San Diego, CA: Academic Press.

Paredes-Olay, M. C., & Rosas, J. M. (1999). Within-subjects extinction and renewal in predictive judgments. *Psicologica, 20,* 195–210.

Pavlov, I. P. (1927). *Conditioned reflexes*. London: Clarendon Press.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.

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

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 37*(B), 1–21.

Shanks, D. R. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology, 84,* 319–354.

Shanks, D. R., López, F. J., Darby, R. J., & Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Causal learning* (Vol. 34, pp. 265–311). San Diego, CA: Academic Press.

Vila, N. J. (2000). Extinción e inhibición en juicios de causalidad [Extinction and inhibition in causality judgments]. *Psicologica, 21,* 257–273.

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

Wasserman, E. A., Kao, S.-F., Van Hamme, L. J., Katagari, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Causal learning* (Vol. 34, pp. 207–264). San Diego, CA: Academic Press.

Received January 4, 2001
Revision received February 1, 2002
Accepted February 1, 2002 ■