

Frequency of judgment as a context-like determinant of predictive judgments

MIGUEL A. VADILLO, SONIA VEGAS, and HELENA MATUTE
Universidad de Deusto, Bilbao, Spain

Several studies have shown that predictive and causal judgments vary depending on whether the question used to assess the relationship between events is presented after each piece of information or only after all the available information has been observed. This effect could be understood by assuming that in the two cases people perceive that the test question requires that different sets of evidence be taken into account. This hypothesis is tested in the present experiments through contextual manipulations that take place at the time of training and at the time of test. Our results show that people use this contextual information to infer which set of events should be considered when making their subjective assessments. The results are at odds with current theoretical approaches, but it is possible to develop mechanisms that would allow these models to account for the observed evidence.

Learning to predict future events from present events is one of the most powerful adaptive tools, since it allows an organism to find the necessary resources for survival and to avoid dangerous situations. Given its importance, this kind of predictive learning was the central focus of animal behavior research throughout the twentieth century. During the last decades, predictive learning has also become important in the area of human cognition, where it has given rise to a great amount of empirical and theoretical research.

The vast amount of evidence provided by this research has sometimes turned out to be quite difficult to explain by the available theoretical approaches. Many variables usually neglected by theoretical models influence the process of human learning of predictive relations among events or the way in which humans use the acquired information. Among other things, it has been shown that the probe question used to assess participants' judgment (Matute, Arcediano, & Miller, 1996; Matute, Vegas, & De Marez, 2002; but see also Cobos, Caño, López, Luque, & Almaraz, 2000), the perceived properties of the cue and the outcome (De Houwer, Beckers, & Glautier, 2002; Mitchell & Lovibond, 2002), and the frequency with which judgments are emitted (Catena, Maldonado, & Cándido, 1998; Collins & Shanks, 2002; Matute et al., 2002) can be responsible for the differential results observed.

Several experiments have shown that predictive and causal judgments—that is, the subjective assessment of the relationship between two events—can be altered depending on the frequency with which judgments are produced (see, e.g., Catena et al., 1998; Collins & Shanks, 2002; Matute et al., 2002; but also see Wasserman, Kao, Van Hamme, Katagari, & Young, 1996, who observed the same result regardless of whether participants responded in every trial or only at the end of training). Consider, for example, a situation in which participants receive a series of cue–outcome trials followed by a series of trials in which the cue is no longer followed by the outcome. If participants are required to make their judgments while they are still receiving the information, their judgments will be strongly influenced by the most recent information received (i.e., recency effects will be observed). In contrast, if participants must make their judgments only after seeing all the information, their evaluations will match the overall statistical relation between the events quite well. That is, they will tend to integrate (or average) the information received during the two phases. In some studies (e.g., Dennis & Ahn, 2001), a primacy bias (i.e., a tendency to give more importance to the evidence presented in the first phase) has even been observed in this situation.

This response mode effect is important because it cannot be easily explained by most theoretical models of causal and predictive learning. On the one hand, associative models (e.g., Dickinson & Burke, 1996; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994) assume that the learning process consists in the progressive strengthening or weakening of the association between the mental representation of the cue and the outcome. According to these models, the predictive value of a cue is updated after every trial. This updating process accounts quite well for the recency effect observed in experiments using a trial-by-trial response mode and, as-

Support for this research was provided by Grant PI-2000-12 from Departamento de Educación, Universidades, e Investigación of the Basque Government to H.M. M.A.V. and S.V. were supported by F.P.I. Fellowships (BFI01.31 and BFI00.138, respectively) from the Basque Government. We thank Lorraine Allan, Leyre Castro, Jan De Houwer, and Gabriel A. Radvansky for their helpful comments on an earlier draft, and Fernando Blanco and Ann Meulders for their assistance in conducting the third experiment. Correspondence concerning this article should be addressed to H. Matute, Departamento de Psicología, Universidad de Deusto, Apartado 1, 48080 Bilbao, Spain (e-mail: matute@fice.deusto.es).

suming certain parameter values, it could also account for the accurate judgments obtained when a global response mode is used. But it is hard to see why these parameters should vary depending on the response mode.

On the other hand, statistical models view the process of predictive learning as the computation of the contingency between the cue and the outcome (Allan, 1980; Cheng, 1997; Cheng & Novick, 1992; Jenkins & Ward, 1965). According to this perspective, people store all the relevant information presented during the two training phases and, when asked to rate the relationship between the cue and the outcome, they simply use this stored information to compute the relevant statistical measure. The predictions of these models match quite well the results obtained when a global response mode is used. However, they cannot account for the recency effects obtained with a trial-by-trial procedure, unless it is assumed that the evidence presented during the first phase of training is neglected in this case (see Matute et al., 2002).

A potential explanation for the divergent results obtained when the response mode is manipulated is provided by belief revision models, which state that the information available is processed and stored differently in each case (Catena et al., 1998; Collins & Shanks, 2002; Hogarth & Einhorn, 1992; Pennington & Hastie, 1992). According to these models, each judgment is produced taking into account the evidence presented between the last judgment and the present judgment. This evidence is used to compute the strength of the relationship between the cue and the outcome and to update the assessment made in the last judgment accordingly. Once a judgment is made, all previous information is lost. When only one judgment is required, this procedure leads to an integrative judgment, since all the information will be considered. By contrast, when several judgments are required, this procedure predicts that older information will be lost as a function of the updating process and, thus, the judgment will be highly influenced by the most recent information received.

Even though this perspective provides a better account than most statistical and associative models, it also has some limitations. For example, Matute et al. (2002, Experiments 1A–1C) showed that not only the response mode but also the type of question used (e.g., predictive vs. causal) could yield quite different results. If the recency effect obtained with trial-by-trial procedures is due to a particular information-processing and -storing strategy, it is not clear why this effect should disappear as a function of test question. In a different experiment, Matute et al. (2002, Experiment 4) also observed that the recency effect typically observed in trial-by-trial groups can be counteracted if participants are exposed to an instructional screen just between the last training trial and the test trial. After these instructions, an integrative judgment, rather than a recency-based one, can be observed. This result implies that even if participants have been exposed to a trial-by-trial procedure, they have access to all the information received, and not just to the in-

formation received between the last and the present judgments, as suggested by belief revision models.

According to Matute et al. (2002), people do store all the information received, and the divergent results obtained using trial-by-trial and global response modes occur because of the different interpretation of test demands that participants make in each case. When participants are asked only to make a single assessment at the end of training, by default they interpret that this question refers to all the information presented during training and, thus, they make an integrative estimation. On the other hand, when a judgment is required after each piece of information is presented, people tend to perceive that they are being asked to assess the current status of the cue, which makes them focus on the most recent trials.

The advantage of this explanation is that it accounts not only for the response mode effect, but also for the effect of other testing variables, such as the type of question or the role of instructions, by the very same mechanism (Matute et al., 2002). Different probe questions can give rise to different results under a trial-by-trial response mode because some questions (e.g., predictive questions) make participants think that the most recent information is most relevant, whereas others (e.g., causal questions) suggest that all the information should be taken into account. As an example, when contingencies change over time and people have to make frequent predictions about the expected occurrence or nonoccurrence of the outcome, they tend to make predictions that correspond to the current state of affairs; it would not make much sense to make predictions on the basis of how things used to be in the past. By contrast, when people are asked to rate the causal relationship between the cue and the outcome, it would not make much sense to use only the most recent information, since causal relationships are supposed to be relatively constant over time; in this case people tend to give more integrative responses. In the same framework, the effect of an instructional screen presented just before testing becomes straightforward. Depending on what information people are requested to consider at test, they will demonstrate one or the other type of effect, thereby showing that they have not lost the older information.

A testable prediction of this view is that *any* manipulation that affects the perceived demands of the test question will have an impact on the subjective assessments reported. When asked to make a judgment about the predictive relationship between a cue and an outcome on the basis of ambiguous evidence (e.g., when the relation between the cue and the outcome is not stable and changes with time), participants must infer whether they are being asked about the general relationship (in which case they should take into account all the information available) or about any of the momentary relationships (in which case they should take into account only the information provided during the relevant block of trials). Therefore, if participants have access to any type of information that might help them decide which of the different sets of in-

formation should be considered at test, that information will modulate their judgment.

An interesting way in which this hypothesis can be tested is by studying how the manipulation of the response mode interacts with contextual manipulations. Contextual manipulations are often used in animal learning research (for a review, see Bouton, 1993). In order to help nonhuman animals disambiguate the information provided by changing contingencies, researchers in the animal learning tradition have often introduced contextual cues that signal different sets of trials containing contradictory information. These contextual manipulations have recently been shown to affect human causal and predictive judgments as well (Matute & Pineño, 1998; Pineño & Matute, 2000; Vila & Rosas, 2001).

According to our hypothesis, participants responding on a trial-by-trial basis will no longer make their default recency-based estimation if the contextual cues presented at test suggest that earlier information is more relevant than recent information. Similarly, the integrative or intermediate judgments usually observed in global procedures will not be observed when the contextual cues presented at test indicate that one set of trials is more relevant than the other. Thus, these experiments will test whether the information provided by contextual cues can counteract the response mode effect.

EXPERIMENT 1

Matute et al. (2002) suggested that, when a trial-by-trial response mode is present, recency effects appear because people perceive that the final test question is just one more instance of the same question, identical to that presented in the preceding trials. Because nothing in this question indicates that it is different from the previous ones or that it refers to the evidence in its entirety, people assume it refers to the most recent block of evidence. If this is true, it follows that if something changed at the time of test, people would notice that the test question was different from the previous ones—in other words, that it does not refer only to the information most recently presented. In this case, trial-by-trial participants could be made to make an integrative judgment instead of their default recency-based response.

This prediction is tested in Experiment 1, in which three groups of participants received the same cue–outcome pairings but made their estimations in different contexts and with different response modes. Two of these groups, Global-111 and TT-111, were the standard global and trial-by-trial groups, respectively. For these groups, the context remained unchanged during two training phases and a test phase. We use the term “111” to indicate that the context was the same in these three phases. These groups should replicate the basic response mode effect (see, e.g., Matute et al., 2002). The third group, TT-112, was a trial-by-trial group for which the test phase took place in a novel context, different from that of the two previous training phases. The term “112” is used to in-

dicating that the context is the same for the two training phases but different for the test phase. If our hypothesis is correct, this group should behave at test in a way similar to that of the global group, rather than show the recency bias that is usually observed in the standard trial-by-trial group.

Method

Participants and Apparatus. Fifty-four undergraduate students from Deusto University volunteered for the experiment. Random assignment resulted in 17 participants in Group TT-111, 18 participants in Group TT-112, and 19 participants in Group Global-111. All the participants performed the experiment at the same time in a large computer room. Adjacent participants were seated about 1.5 m apart from each other and were exposed to different experimental conditions.

Design. The design summary is shown in Table 1. During training, all groups of participants received two blocks of trials. During the first block, they were exposed to 20 trials in which the cue was always followed by the outcome (C–O). They then saw another 20 trials in which the cue was no longer followed by the outcome (C–noO). The only differences between the three groups were the response mode and the contextual manipulations at the time of test. Two of the groups responded in a trial-by-trial mode. For one of them, all three phases took place in Context 1 (Group TT-111); for the other, Phases 1 and 2 took place in Context 1 and the test phase took place in a second context (Group TT-112). The third group was a global group for which the three phases took place in the same context (Group Global-111).

Procedure. The experimental task was a computerized version of the standard allergy task used in experiments on causal and predictive learning (see, e.g., Wasserman, 1990). The participants were shown the medical records of fictitious patients. The record of each patient consisted of two consecutive cards. The cue card was presented first in each trial and indicated whether or not that patient had taken a fictitious medicine called Dugetil. The outcome card then indicated whether or not the patient had developed an allergic reaction to the medicine. Contextual cues were text labels that were presented on each card and given symbolic contextual value through instructions, as is usually the case when such manipulations are performed with humans (see, e.g., Vila & Rosas, 2001). The two contexts used in this experiment (i.e., “Fifteenth Century” and “Sixteenth Century” text labels that appeared at the top of the screen) were counterbalanced across participants.¹ The instructions were presented on two consecutive screens. An English translation of the Spanish instructions is offered in Appendix A.

In each trial, the participants from Groups TT-111 and TT-112 were asked to rate the degree to which they thought that the current patient would develop the allergic reaction. This question was always presented on the same card that indicated that the patient had taken

Table 1
Design Summary of Experiment 1

Group	Response Mode	Training		Test
		Phase 1	Phase 2	
Global-111	Global	(C–O)1	(C–noO)1	(C)1
TT-111	Trial by trial	(C–O)1	(C–noO)1	(C)1
TT-112	Trial by trial	(C–O)1	(C–noO)1	(C)2

Note—The cue (C) represents a fictitious medicine presented in every trial; the outcome is an allergic reaction that is present during Phase 1 (O) but not during Phase 2 (noO). The first and second phases of the study always occurred in the same context. The test trial took place in the same context for Groups Global-111 and TT-111 but in a different context for Group TT-112. The two contexts are represented by the number outside the parentheses.

Dugetil. This question can be translated into English as, “To what degree do you think this patient will develop an allergic reaction?” To register their judgments, the participants were asked to introduce a number from 0 to 100 by using the UP and DOWN arrow keys and the ENTER key on the keyboard. The participants in Group Global-111 saw this question in the test trial only. For Groups TT-111 and Global-111, the test trial took place in the same context as the training phases. By contrast, for Group TT-112, this context changed at test.

Results and Discussion

The mean judgments at test are shown in Figure 1. As can be seen in this figure, mean judgments at test were higher in Group Global-111 than in Group TT-111. Judgments at test were also higher in Group TT-112 than in Group TT-111. These impressions were confirmed by a one-way analysis of variance (ANOVA) performed on the judgments at test, which revealed a main effect of group [$F(2,51) = 10.27, MS_e = 1,385.37, p < .001$], and by planned comparisons with α divided by the number of contrasts according to the Bonferroni correction procedure ($\alpha = .05/3 = .017$), which revealed a significant difference between Groups TT-111 and Global-111 [$t(34) = 5.01, p < .001$] and a significant difference between Groups TT-111 and TT-112 [$t(33) = 3.04, p < .01$], but no differences between Groups TT-112 and Global-111 [$t(35) = 1.36, p = .18$].

The difference observed between Groups Global-111 and TT-111 replicates the response mode effect observed by Matute et al. (2002). Most important, the significant difference found between the two trial-by-trial groups suggests that, as was expected, a simple contextual manipulation at the time of test to make participants perceive that the test question might be different from the previous question gives rise to different predictive judgments. Moreover, the absence of differences between Groups Global-111 and TT-112 strongly suggests that, regardless of the response mode, participants can make a statistically accurate judgment whenever the test ques-

tion does not seem to refer only to the last set of trials. This result is consistent with the idea that the recency effect usually observed with trial-by-trial procedures is attributable to the participants’ perceiving that the test trial is identical to the previous one and that the contingencies determining the presence or the absence of the outcome during the last block of trials remain unchanged.

EXPERIMENT 2

The results of Experiment 1 suggest that contextual manipulations at the time of test can affect the results, and that this effect might be due to the differential subjective interpretation that the participants make of the test demands. However, this manipulation was used only in order to prevent the test question from being perceived as just one more instance of the question presented during the most recent training trials. Interesting results could also be observed if contextual cues were manipulated not only at the time of test, but also during the training stages. These contextual cues would become more informative if they were used to “mark” the different blocks of information. If during each of these blocks different contextual cues were used, participants could use the presence of these “marking” contextual cues during the test phase to infer which block of trials should be considered most relevant when dealing with the test question. One purpose of Experiment 2 was to test this prediction.

In addition, in Experiment 1 contextual manipulations were used only in a group that had been exposed to a trial-by-trial response mode. Thus, a second aim of Experiment 2 was to investigate whether this kind of contextual manipulation could also have an effect when the response mode is global. Therefore, all the groups in Experiment 2 were global groups—that is, they were asked to make a single judgment after having received all the information.

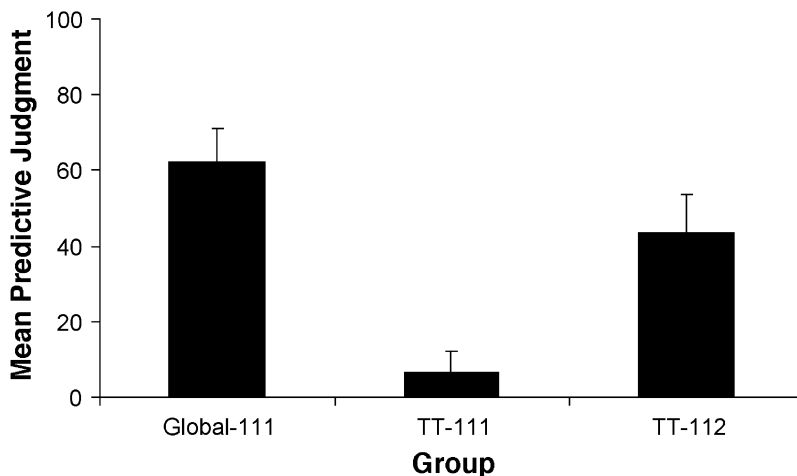


Figure 1. Mean predictive judgment at test for the three groups in Experiment 1. Error bars represent the standard errors of the means.

Method

Participants and Apparatus. Fifty-one undergraduate students from Deusto University volunteered for the study. None of these participants had taken part in the previous experiment or in any related experiment. Random assignment resulted in 13 participants in Group Global-111, 12 participants in Group Global-121, 13 participants in Group Global-112, and 13 participants in Group Global-122. In this experiment, the participants were tested in small groups in a room with individual cubicles. Except for this change, the apparatus was the same as in the previous experiment.

Design and Procedure. In this experiment, the four global groups described in Table 2 were used. As in the previous experiment, all groups of participants received a block of 20 trials in which the cue was always followed by the outcome and a block of 20 trials in which the cue was never followed by the outcome. In this case, the four groups of participants were asked to make a single predictive judgment after having seen the two blocks of trials (i.e., global response mode). The groups differed in relation to the contextual manipulations during the training stages and at the time of test. For Group Global-111, the two training phases and the test phase took place in Context 1. For Group Global-112, the context remained constant during both training phases but the test trial took place in a second context. For Group Global-121, the first training phase and the test phase were conducted in the same context, whereas the second training phase took place in a second context. For Group Global-122, the second training phase and the test phase were performed in the same context, which was different from that of the first training phase. All other procedural aspects were the same as in Experiment 1.

Results and Discussion

The results of this experiment are shown in Figure 2. As the figure shows, the contextual manipulations had a large effect on the predictive judgments. As was expected, when the contextual cues presented at the time of test did not provide information about the relative relevance of each block of trials (i.e., when the contextual cues were kept constant across the three phases [Group Global-111] or when testing took place in a novel context [Group Global-112]), the subjective estimations were close to the middle of the scale—that is, the participants tended to integrate (or average) the contradictory infor-

mation received during the two training phases. Moreover, when the contextual cue presented at the time of test was related to one specific block of trials (i.e., Groups Global-121 and Global-122), the estimations were high or low depending on the block of trials that was being signaled by the test context. These impressions were confirmed by a one-way ANOVA performed on the judgments at test, which showed a main effect of group [$F(3,47) = 15.27, MS_e = 794.51, p < .0001$]. As was expected, planned comparisons with α adjusted according to the Bonferroni correction procedure ($\alpha = .05/2 = .025$) yielded no differences between Groups Global-111 and Global-112 [$t(24) = 0.74, p = .46$], but the observed difference between Groups Global-121 and Global-122 proved to be significant [$t(23) = 8.3, p < .001$].

These results suggest that the tendency to integrate the information usually observed under a global response mode arises only when the cues available at the time of test do not provide information that could be used to infer that one set of trials is more relevant than the other. However, if contextual (or other) cues signal each different block of trials, then the presentation of these cues at the time of test can be used to infer that some trials are more relevant. Thus, even under a global response mode predictive judgments can be based on the most recent or on most distant information, rather than on an integrative strategy, whenever participants are induced to think that this will be the most appropriate response. An interesting result in this experiment is the lack of a significant difference between Groups Global-111 and Global-112. This result contrasts with the clear differences observed in Experiment 1 between Groups TT-111 and TT-112 and suggests that the high response observed in the 112 groups is not merely due to the fact that testing in a new context elevates the level of responding that would be observed in the corresponding 111 group. If this were the case, then responding in Group Global-112 should also have been higher than responding in Group Global-111. However, in order to properly address the comparison between the global and TT groups of Experiments 1 and 2 when there is and when there is not a context shift, it would be necessary to perform an additional experiment in which all these groups are included.

EXPERIMENT 3

The results presented up to this point show that contextual cues can play a key role when participants deal with a predictive task based on conflicting evidence. Moreover, the effect of these contextual cues seems to overcome the influence of the response mode effect. Participants working under trial-by-trial procedures no longer make judgments based only on recent information if the test context suggests that the present trial does not belong to the last set of trials (Experiment 1). Similarly, participants required to make a single (i.e., global) judgment at the end of training no longer make their default integrative estimations if contextual cues suggest that a partial estimation is more ade-

Table 2
Design Summary of Experiments 2 and 3

Group	Response Mode	Training		Test
		Phase 1	Phase 2	
Global-111	Global	(C–O)1	(C–noO)1	(C)1
Global-121	Global	(C–O)1	(C–noO)2	(C)1
Global-112	Global	(C–O)1	(C–noO)1	(C)2
Global-122	Global	(C–O)1	(C–noO)2	(C)2
TT-111	Trial by trial	(C–O)1	(C–noO)1	(C)1
TT-121	Trial by trial	(C–O)1	(C–noO)2	(C)1
TT-112	Trial by trial	(C–O)1	(C–noO)1	(C)2
TT-122	Trial by trial	(C–O)1	(C–noO)2	(C)2

Note—In Experiment 2, only the global groups were tested. In Experiment 3, all eight groups were tested. The cue (C) represents a fictitious medicine presented in every trial; the outcome is an allergic reaction present during Phase 1 (O) but not during Phase 2 (noO). In these experiments, the different phases of the study could occur either in Context 1 or in Context 2 (represented by the numbers outside the parentheses). The groups differed with regard to whether the second training phase and/or the test phase took place in Context 1 or in Context 2.

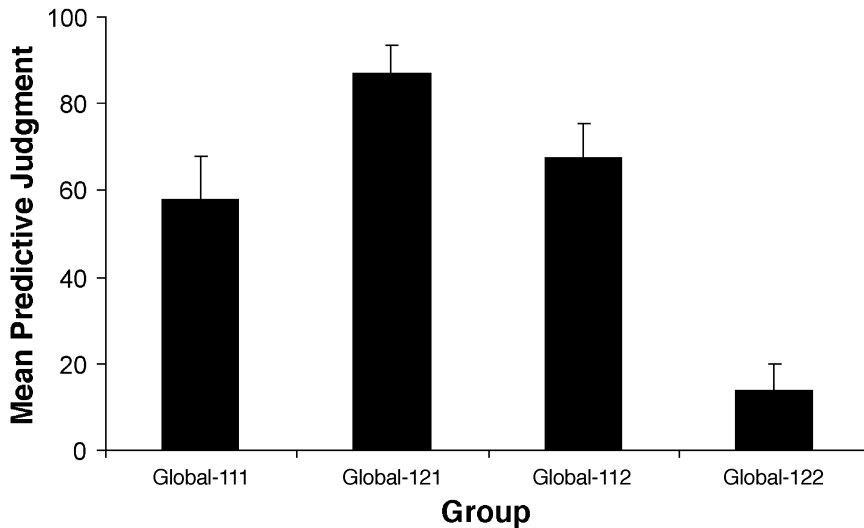


Figure 2. Mean predictive judgment at test for the four groups in Experiment 2. Error bars represent the standard errors of the means.

quate (Experiment 2). Thus, our next step was to compare within a single experiment the way these manipulations affect both trial-by-trial and global procedures. In order to allow for such comparisons, Experiment 3 comprises all the groups from Experiment 2 together with analogous trial-by-trial groups. This will allow us to explore other predicted similarities between certain effects of the response mode and certain contextual manipulations that have not yet been tested. In addition, contexts were manipulated in a different way so as to make sure that the results of Experiments 1 and 2 can be generalized when different contextual cues are used to retrieve the information related to one or the other set of trials.

Method

Participants and Apparatus. One hundred ninety-one undergraduate students from Deusto University volunteered for the study. None of the participants had taken part in any related experiment. Half of the participants were tested in a large computer room as in Experiment 1, whereas the other half were tested in individual cubicles as in Experiment 2. In both cases, the participants were randomly distributed across the eight experimental groups.

Design and Procedure. The design of this experiment is summarized in Table 2. Four groups were exposed to a global response mode, whereas four other groups were exposed to a trial-by-trial response mode. Except for the changes in the contextual manipulations mentioned below, all procedural details were the same as in Experiments 1 and 2.

The same allergy task used in the previous experiments was used in this experiment, but a few modifications were introduced in order to improve the contextual manipulations. In Experiments 1 and 2, the contexts were two different centuries in which the fictitious patients were supposed to have lived. This kind of context could be quite strange or unnatural for participants, since it is not clear why the centuries should have an influence on the medicine's effect. Because of this, in Experiment 3 we decided to use a more natural set of fictitious contexts, in a way similar to that used by other human learning researchers (e.g., Vila & Rosas, 2001). In Experiment 3, the participants were informed that the medicines could

have been developed in different laboratories. The laboratory that had produced the medicine was always indicated at the top of the screen. They were named "Laboratory A" and "Laboratory B," and the distinction between these two laboratories was accentuated by the use of different colors (yellow and blue, respectively) for the screen. Both the names and the colors of the two laboratories were counterbalanced. The participants were encouraged to pay attention not only to the medicines and allergies, but also to the laboratories. The instructions for the experiment were modified accordingly by a change to the second screen of instructions. A translation of the second screen of instructions used in this experiment is offered in Appendix B.

Results and Discussion

The mean judgments at test are shown in Figure 3. As Figure 3 shows, both contextual manipulation and response mode had an effect on predictive judgments at test. A 2 (response mode: TT vs. global) \times 4 (contextual manipulation: 111 vs. 121 vs. 112 vs. 122) ANOVA on the final judgments during test yielded a main effect of response mode [$F(1,183) = 20.42, MS_e = 584.03, p < .001$] and a main effect of contextual manipulation [$F(3,183) = 85.56, MS_e = 584.03, p < .001$]. Moreover, the interaction between them was also significant [$F(3,183) = 3.35, MS_e = 584.03, p < .05$]. Five subsequent contrasts with α adjusted according to the Bonferroni correction procedure ($\alpha = .05/5 = .01$) were performed to look for the source of this interaction.

The mean judgments made at test by the participants in Groups Global-111 and TT-111 indicate that when the contextual cues were kept constant during both training and testing the results varied depending on the response mode (see Figure 3). Contrasts showed that judgments at test were lower for Group TT-111 than for Group Global-111 [$t(46) = 4.74, p < .001$], which replicates the basic response mode effect that we also observed in Experiment 1 of the present work.

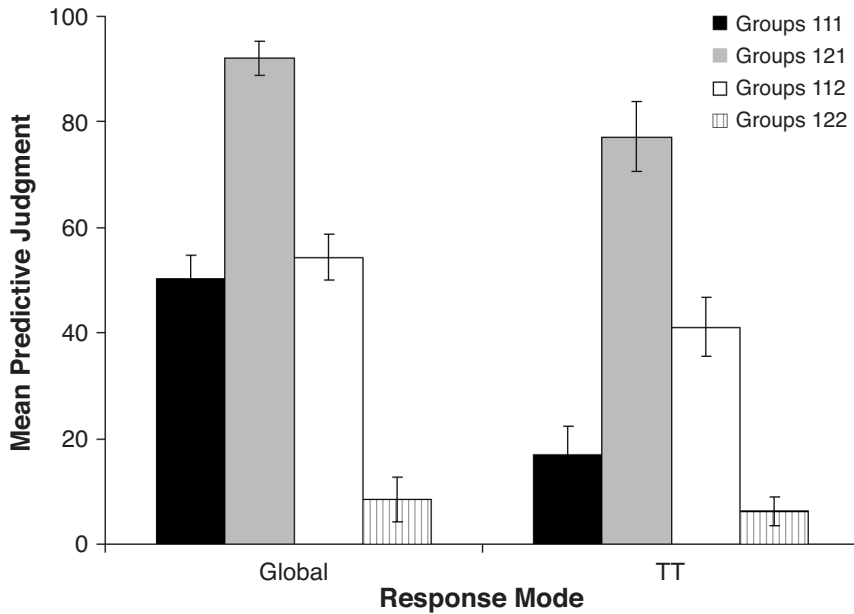


Figure 3. Mean predictive judgment at test for the eight groups in Experiment 3. Error bars represent the standard errors of the means.

Figure 3 also shows that when the contextual cue presented at the time of test signaled one and only one block of trials the final judgment was coherent with the information provided within that block. This occurred regardless of the response mode with which the participants emitted their judgments. Planned comparisons revealed significant differences between Groups Global-121 and Global-122 [$t(46) = 15.76, p < .001$] and also between Groups TT-121 and TT-122 [$t(46) = 8.84, p < .001$]. This suggests that, regardless of their response mode, the participants used the contextual cues presented at the time of test to make their judgments if these cues had had a disambiguating role during training.

Moreover, when the context did not change between the first and the second training phases, making the test judgment in a novel context had an effect on judgments under a trial-by-trial response mode (TT-112) but not under a global response mode (Global-112). Planned comparisons revealed significant differences between TT-111 and TT-112 [$t(47) = 3.04, p < .005$], whereas no differences were observed between Groups Global-111 and Global-112 [$t(44) = 0.65, p = .52$]. As was expected, the global participants maintained their default integrative judgments, but trial-by-trial participants, whose default response would be a recency response, opted for an integrative response when tested in a novel context. In both cases, the result was an integrative judgment when the test context was new.

GENERAL DISCUSSION

Although some of the results presented here can be partially explained by different theoretical models, in their

present state none of the main theoretical approaches can provide a fully satisfactory explanation for all the results observed. As was mentioned in the introduction, the two main theoretical frameworks—namely, the associative framework and the statistical framework—cannot explain the differences obtained when response mode is manipulated. The mechanisms proposed by the associative view can account for the recency effects observed by default under trial-by-trial procedures but not for the integrative judgments generally observed with global procedures. On the other hand, the kind of probabilistic rules proposed by the statistical models closely matches the results generally obtained with global procedures, but it does not explain the tendency to recency observed under a trial-by-trial response mode. Similarly, it is not clear how these models would explain the effects of some of the contextual manipulations used here.

Some models that we have not yet considered could provide partially satisfactory explanations for some of these results. For example, the finding that predictive judgments at test when a specific context is present tend to match the particular cue–outcome contingency signaled by that context (e.g., high ratings in the 121 groups and low ratings in the 122 groups) can be easily explained by any associative model that assumes that what is associated with the outcome is not the particular cue but the cue–context configuration (Pearce, 1987, 1994). According to this model, the participants in the 121 groups should give high ratings because the Cue + Context 1 node, which is activated at test, is associated with the outcome. Similarly, a low rating in the 122 groups is expected because the Cue + Context 2 presented at test is associated with the absence of the outcome. However,

this model cannot account for the basic difference found between the trial-by-trial groups and the global groups (e.g., TT-111 vs. Global-111), which is one of the basic phenomena observed in these experiments.

In a similar fashion, certain versions of statistical models could be used to explain some of our results. For example, one of our major findings is that participants responding on a trial-by-trial basis stop giving low ratings at test if the test takes place in a different and new context (e.g., TT-112). According to the probabilistic contrast model (Cheng & Holyoak, 1995), this higher rating might be due to the uncertainty people might experience when they are asked to give a rating referring to an unknown context. During training, the participants in Group TT-112 had received information about the probability of the outcome given the cue in the presence of one context. When the new context is presented, they cannot be sure that the cue will be followed by the outcome in this new context, and so they make an intermediate judgment. However, this model remains unable to account for the recency-based judgments observed in Group TT-111.

In order to account for these phenomena, a suitable theoretical model would have to allow for the independent storage and retrieval of the information presented during the different stages. Most current associative models fail to meet this requirement, since they postulate that after a participant has seen all the evidence a single associative strength is stored. Nor can current statistical models meet this criterion, given that according to them all the information available will be used to compute the appropriate contingency measure; whether this information comes from stable or unstable evidence lacks importance. Nevertheless, to overcome this limitation both associative and statistical models can be extended with a few assumptions.

Extending Associative Models

One important limitation of the most popular associative models (see, e.g., Rescorla & Wagner, 1972) is that they allow for the storage of only a single associative strength, which summarizes the perceived relationship between the cue and the outcome. When the associative mechanism faces the problem that the contingencies are unstable, its updating rule will adjust the connections according to the present contingencies and, in this process, the associative strength developed for previous contingencies will be lost. This effect is known as *catastrophic forgetting* (Lewandowsky, 1991; McCloskey & Cohen, 1989; Ratcliff, 1990). Although this forgetting mechanism accurately predicts the recency effects obtained in some of the groups in the present experiments, it also prevents associative models from accounting for any result in which the evidence received first is used at test. However, this feature does not affect all associative models in the same way.

The design used in the experiments presented here (i.e., a cue always followed by an outcome in the first

phase and never followed by that outcome in the second phase) is an acquisition–extinction design and has a long tradition in animal conditioning research. The results of many experiments in which this design has been used have provided interesting data that show that associations between a cue and an outcome are not destroyed when the cue no longer predicts the outcome. For example, many studies have shown that the conditioned response can be recovered after extinction if the test trial takes place in a context different from that of the extinction phase (an effect known as *renewal*; Bouton & Bolles, 1979) or if there is a time interval between the end of the extinction phase and the test phase (*spontaneous recovery*; Pavlov, 1927). Moreover, these effects have also been replicated in human contingency learning experiments (see, e.g., Pineño & Matute, 2000; Rosas, Vila, Lugo, & López, 2001). The evidence provided by these studies has motivated the development of new associative models that are less sensitive to catastrophic forgetting.

According to some associative models (e.g., Bouton, 1993, 1997; Pavlov, 1927), the information relative to positive and negative contingencies is stored in independent associations (excitatory and inhibitory, respectively). Thus, when the contingency between the cue and the outcome changes from positive to negative, a new inhibitory association is developed which does not require the destruction of the previous excitatory one. According to Bouton (1993, 1997), the problem of storing these two associations is that the cue becomes an ambivalent predictor of the outcome. In this situation, the factor that determines which of the two associations is expressed is the relative activation of each of them during testing, which depends on the physical or temporal contextual cues that are available at test. If the test phase is presented immediately after the extinction stage (i.e., in the same temporal context) or in the same physical context, then the association developed second (i.e., the inhibitory one) will be expressed; but if the test phase is conducted at a different time or in a context different than that of the second stage, then the association developed first will be expressed.²

Within this framework, the different results of the experiments presented here can be understood as involving the kind of context-based retrieval of the associations proposed by Bouton (1993, 1997) to account for the renewal effect and other animal conditioning phenomena. When different contexts are present during each training phase (e.g., Groups 121 and 122 in Experiment 3), the presence of one or the other context at the time of test will determine which association will be expressed. The only extension needed in this type of model is the assumption that when the context is constant the second association will have a higher activation in memory at the time of test (Group TT-111), *unless* there is anything in the test question that makes people think that the test trial is different from the most recent trials. That is, not only contexts but also other types of variables can be used to separate the test trial from the most recent trials and, in consequence, reactivate older information in mem-

ory. Instances of variables that produce this rupture between the second phase and the test trial are the presentation of a new question at the time of test (Group Global-111) and the presentation of the same question in a new context (Group TT-112). In these cases, both the first and the second acquired associations will be partially activated by the novel test context or question.

Extending Statistical Models

Similarly, current statistical models also face important limitations in accounting for the present results. They would need to allow for the separation of the different sets of evidence and for a relatively independent storage of the information contained in them. Although current statistical models do not accomplish this requirement directly, some recent normative considerations regarding contextual and conjunctive causal power (Cheng, 2000) could supply the theoretical framework to overcome this problem.³ According to Cheng's (2000) normative analysis, the power of a given cause can change from one context to another and, when this is the case, the assessments of the relation between the cause and the effect cannot be generalized from one context to another. Thus, in the groups in which we used contextual switches, our participants may have regarded the different contextual cues as reflecting different contextual causal powers. From this perspective, the independent use of the information contained at each stage would be perfectly coherent. In this case, the judgment made at the time of test will depend on the contextual cues then presented. However, the results observed when the contextual cues are kept unchanged during the two training phases (i.e., when the contexts were only temporal, such as in Group TT-111) are more difficult to explain. Although the information participants receive can also be regarded as reflecting different contextual causal powers, the relevant contextual factors are unobservable and their presence or absence has to be inferred.

Thus, as was the case of most associative models, the main limitation of statistical models, in their present state, is that they cannot compute contingency values separately for each block of evidence. In this case, when asked to make an assessment the participant would just use all the evidence available, regardless of its distribution over time. When a different contextual cue is presented at different phases (Groups 121 and 122 in these experiments), this contextual cue might be used to separate the information contained in each phase. Nevertheless, this separation cannot be directly performed when these contextual cues are absent (e.g., Groups TT-111, TT-112, Global-111, and Global-112 in these experiments). Of course, a partition of the available information can be established a posteriori by researchers, but no a priori mechanisms have yet been suggested to accomplish this task. However, the development of such mechanisms would not be difficult.

In order to account for such results, a statistical model would need the following extension: a mechanism to detect, in the absence of contextual cues, whether the evi-

dence presented reflects a single homogeneous contingency or, by contrast, a variable interevent relation. In the second case, the model should also be able to identify the different blocks in which the available information could be optimally separated. Both these tasks could be accomplished by an algorithm that tests the probabilistic coherence of the sequence of events under analysis. In our experiments, for example, the evidence presented taken as a whole suggests that the probability of the outcome given the cue is .50. However, under this general conditional probability it is highly improbable that in a sequence of 20 consecutive trials all of them would consist in the cue followed by the outcome (Phase 1 in our experiments) or in the cue followed by no outcome (Phase 2 in our experiments). This suggests that a momentary probability of the outcome given the cue equal to 1, rather than a general conditional probability of .50, might be responsible for the sequence of events in Phase 1, and that a momentary probability of 0 might be responsible for the Phase 2 sequence. Accordingly, the evidence can be better seen as reflecting two different phases with different interevent relations. The sequences of evidence that are maximally improbable given the conditional probabilities suggested by the information in its entirety might be regarded as independent blocks. The participants' judgments could be based on either of the different blocks or on the conjunction of all the information presented, depending on the perceived similarity between the situation of the test trial and that of each of the training phases.

In our experiments, when each set of training trials is presented together with particular contextual cues (i.e., Groups 121 and 122), the contextual cues presented at the time of test will provide the information required to determine which of the partial blocks of information is more applicable. On the other hand, when these contextual cues do not provide such information (i.e., Groups 111 and 112), since it is not clear which of the partial contingencies is more adequate, the judgment will be made taking into account both blocks of information, unless the test question itself indicates which is more relevant. For example, for the participants in Group Global-111, the test question is completely new and, although nothing else changes, this signals that this question does not belong to the most recent set of trials. In this case, an integrative judgment is emitted. However, for the participants in Group TT-111 nothing changes between the last training trial and the test trial, and this leads the participants to assume that the particular contingency of the second phase still applies. As can be observed, once the different blocks of trials have been separated, the factors that could determine which of the partial contingencies should be taken into account at the time of test would be quite similar to those described in the discussion of the extension of associative models above.

Some Concluding Comments

It is worth noting that the type of extensions we have suggested for the main theoretical approaches will make

them progressively more indistinguishable from each other and their differential predictions harder to test. However, they need not be seen as competing theoretical alternatives but, rather, as different levels of explanation, one of them (the statistical account) being a computational or general approach and the other (the associative account) an algorithmic description of the low-level mechanisms involved in causal and predictive learning.

Of course, the kind of account that we have suggested can also be extended to belief revision models (Catena et al., 1998; Hogarth & Einhorn, 1992), provided they allow for the storing of all the information regardless of the response mode. These models should explicitly assume that even under trial-by-trial responding some kind of record of all the information presented needs to be stored. Moreover, these models should assume some mechanism by which the different blocks of trials could be separated. These procedures would allow for the same type of extensions as those that we have proposed for associative and statistical models.

These experiments, together with those of Matute et al. (2002), clearly show that the most well-known theoretical approaches for the study of causal and predictive induction are, in their present state, unable to account for very simple experimental results. However, a model's lack of fitness need not necessarily mean that it is wrong; it could also mean that it is just incomplete. As we have tried to show, the major theoretical approaches can easily be extended in order to take into account many of the neglected variables that obviously influence the predictions that people make. Thinking about the limitations of several theoretical models and about the means by which these limitations might be overcome in most models could lead us to a deeper understanding of the processes underlying human causal and predictive judgment.

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.
- BOUTON, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian conditioning. *Psychological Bulletin*, **114**, 80-99.
- BOUTON, M. E. (1997). Signals for whether versus when an event will occur. In M. E. Bouton & M. S. Fanselow (Eds.), *Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles* (pp. 385-409). Washington, DC: American Psychological Association.
- BOUTON, M. E., & BOLLES, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning & Motivation*, **10**, 445-466.
- 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 & Performance*, **24**, 481-495.
- CHENG, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, **104**, 367-405.
- CHENG, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive causal power. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227-253). Cambridge, MA: MIT Press.
- CHENG, P. W., & HOLYOAK, K. J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contingency and prediction. In H. L. Roitblat & J.-A. Meyer (Eds.), *Comparative approaches to cognitive science* (pp. 271-302). Cambridge, MA: MIT Press.
- CHENG, P. W., & NOVICK, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, **99**, 365-382.
- COBOS, P. L., CAÑO, A., LÓPEZ, F. J., LUQUE, J. L., & ALMARAZ, J. (2000). Does the type of judgement required modulate cue competition? *Quarterly Journal of Experimental Psychology*, **53B**, 193-207.
- COLLINS, D. J., & SHANKS, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, **30**, 1138-1147.
- DE HOUWER, J., BECKERS, T., & GLAUTIER, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, **55A**, 965-985.
- DENNIS, M. J., & AHN, W.-K. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition*, **29**, 152-164.
- DICKINSON, A., & BURKE, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, **49B**, 60-80.
- GODDARD, M. J. (1999). Renewal to the signal value of an unconditioned stimulus. *Learning & Motivation*, **30**, 15-34.
- HOGARTH, R. M., & EINHORN, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, **24**, 1-55.
- JENKINS, H., & WARD, W. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, **7**, 1-17.
- LEWANDOWSKY, S. (1991). Gradual unlearning and catastrophic interference: A comparison of distributed architectures. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 445-476). Hillsdale, NJ: Erlbaum.
- 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., & PINEÑO, O. (1998). Cue competition in the absence of compound training: Its relation to paradigms of interference between outcomes. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 38, pp. 45-81). San Diego: Academic Press.
- 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.
- 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: Academic Press.
- MITCHELL, C. J., & LOVIBOND, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *Quarterly Journal of Experimental Psychology*, **55B**, 311-329.
- NAKAJIMA, S., TANAKA, S., URUSHIHARA, K., & IMADA, H. (2000). Renewal of extinguished lever-press responses upon return to the training context. *Learning & Motivation*, **31**, 416-431.
- PAVLOV, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London: Oxford University Press.
- PEARCE, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, **94**, 61-73.
- PEARCE, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, **101**, 587-607.
- PENNINGTON, N., & HASTIE, R. (1992). Explaining the evidence: Tests of story model for juror decision making. *Journal of Personality & Social Psychology*, **62**, 189-206.
- PINEÑO, O., & MATUTE, H. (2000). Interference in human predictive learning when associations share a common element. *International Journal of Comparative Psychology*, **13**, 16-33.
- 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 non-

reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

- ROSAS, J. M., VILA, N. J., LUGO, M., & LÓPEZ, L. (2001). Combined effect of context change and retention interval on interference in causality judgments. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 153-164.
- TAMAI, N., & NAKAJIMA, S. (2000). Renewal of formerly conditioned fear in rats after extensive extinction training. *International Journal of Comparative Psychology*, *13*, 137-146.
- 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.
- VILA, N. J., & ROSAS, J. M. (2001). Reinstatement of acquisition performance by the presentation of the outcome after extinction in causality judgments. *Behavioural Processes*, *56*, 147-154.
- 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: Vol. 34. Causal learning* (pp. 207-264). San Diego: Academic Press.

NOTES

1. By using different centuries as our symbolic contexts, we do not mean to extract any particular conclusions with respect to the role of real temporal contexts. The only point we mean to make is that contextual cues (of one type or another) associated with one or the other block of trials can counteract the response mode effect. In Experiment 3, we use (perhaps more natural) symbolic physical contexts.

2. It should be noted that in the present experiments (see also Matute et al., 2002) as well as in several experiments in the animal literature (see, e.g., Goddard, 1999; Nakajima, Tanaka, Urushihara, & Imada, 2000; Tamai & Nakajima, 2000), the results show that when testing takes place in a novel context, the expression of the excitatory association is not complete and is usually weaker than when it is tested in the context in which it was acquired. This suggests that when the test context is novel, the two associations, and not only the one acquired first, are partially active in memory. This is a slight deviation from Bouton's (1993) predictions, but it does not weaken the extension of this theory that we are suggesting to account for human predictive learning.

3. Although Cheng's (2000) analysis applies theoretically only to causal relations and our task requires just predictive judgments (whose normative referents are different), the causal scenario used in these experiments could induce participants to make their predictive judgments on the basis of the underlying causal structure.

APPENDIX A

Translation of Instructions for Experiments 1 and 2

Screen 1

Imagine that you are an allergist who wants to study to what degree the consumption of a medicine called Dugetil causes, as a side effect, an allergic reaction. The medical records of a series of patients will be presented. Based on them, you will have to 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 or not the patient developed 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 a particular patient is going to develop the allergic reaction.

Screen 2

But not everything will happen in the same period of time! You will have the opportunity to travel through time with a newly developed time machine. Thus, you will be able to see the cards of patients living in different centuries. At the top of the screen you will see the century you are in. Pay attention to this information, since it is important, but don't forget that the medical records are equally important.

At certain points, the time machine will automatically transport you to a century that may or may not be the same as the one you are in. The machine will warn you when this process is taking place.

APPENDIX B

Second Screen of Instructions for Experiment 3

But be careful, because the medicines taken by the patients might have been developed in different laboratories. That is, you will see the medical records of patients that have taken medicines from different laboratories. At the top of the screen you will always see which laboratory is involved. Pay attention to that message because it is important, but be aware that the medical records are equally important.