Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 3, 672-694

0278-7393/98/$3.00

Effects of Trial Order on Contingency Judgments: A Comparison of Associative and Probabilistic Contrast Accounts

Francisco J. López
Universidad de Málaga

David R. Shanks
University College London

Julián Almaraz and Pablo Fernández
Universidad de Málaga

Four experiments examined trial sequencing effects in human contingency judgment. In Experiments 1–3, ratings of contingency between a target cue and outcome were affected by the presentation order of a series of trials distributed in 2 distinct blocks and showed a recency bias. Experiment 4 replicated this effect when the trials were partly intermixed. These recency effects are predicted by an associative learning model that computes associative strengths trial by trial and incorporates configural coding of cues but are problematic for probabilistic contrast accounts, which currently have no provision in the contingency computation for the differential weighting of trials as a function of their order of presentation.

Humans and other organisms are very adept at judging whether and to what extent two events are related. People can learn that pressing a light switch causes a light to come on, that eating a particular food makes them ill, that car accidents are more likely on wet roads, and a host of similar event relationships. A great deal of research interest has been recently dedicated to the study of how such judgments are made. In this article, we evaluate a theory of contingency judgment, recently formulated by Cheng and Holyoak (1995), which proposes that individuals compute event relationships in a normative fashion. In addition, the predictions made by this contingency theory are contrasted with those made by a completely different account of how event relationships are detected, namely an associative account (Dickinson, Shanks, & Evenden, 1984; Gluck & Bower, 1988; Kruschke, 1992; Pearce, 1994; Rescorla & Wagner, 1972; Wasserman, Elek, Chatlosh, & Baker, 1993).

Suppose we wish to compute the degree of association between two events that we call a cue (C) and an outcome (O). According to the so-called ΔP rule, the normative metric of event association is given by the difference between the probability of the outcome in the presence of the cue, $P(O/C)$, and its probability in the absence of the cue, $P(O/\overline{C})$:

$$\Delta P = P(O/C) - P(O/\overline{C}). \quad (1)$$

This metric can take on values between 1.0 and −1.0. When the probability of the outcome in the presence of the cue is 1.0 and its probability in the absence of the cue is 0, a connection plainly exists and ΔP is 1.0. For intermediate cases where $P(O/C)$ is greater than $P(O/\overline{C})$, a connection is again present, but in this case an imperfect one. Windows don't always break when hit by balls, but being hit by a ball certainly increases the likelihood that a window will break. When $P(O/C)$ and $P(O/\overline{C})$ are equal, the outcome is no more likely in the presence of the cue than in its absence; obviously, in such cases no relationship exists. When $P(O/C)$ is zero and $P(O/\overline{C})$ is 1.0, the cue perfectly predicts the absence of the outcome. Again, intermediate cases exist when $P(O/C)$ is less than $P(O/\overline{C})$. Here, we would say that a cue–outcome relationship exists, except that it happens to be a negative one.

A number of researchers have suggested that people determine event relationships by maintaining records of the relevant conditional probabilities and then base their contingency judgments on the difference between $P(O/C)$ and $P(O/\overline{C})$ (e.g., Wasserman, 1990). However, the computation of ΔP given in Equation 1 needs to be modified because that equation makes a number of erroneous predictions even in simple situations. For example, consider an experiment by Shanks (1991, Experiment 2), in which participants learned relationships between symptoms (cues) and diseases (outcomes). In a so-called contingent condition, cues A and B were paired together on some trials and reliably predicted outcome O_1 ($AB \rightarrow O_1$). On other trials, B predicted no disease outcome ($B \rightarrow$ no O) and C predicted outcome O_1 ($C \rightarrow O_1$). In a noncontingent condition, presented concur-

Francisco J. López, Julián Almaraz, and Pablo Fernández, Departamento de Psicología Básica, Psicobiología y Metodología, Universidad de Málaga, Málaga, Spain; David R. Shanks, Department of Psychology, University College London, London, England.

The research described here was supported by research grants from the United Kingdom Biotechnology and Biological Sciences Research Council and Economic and Social Research Council (ESRC) and by Dirección General de Investigación Científica y Técnica (PB-91-0028-C03-03) and Junta de Andalucía (HUM0105) from Spain. The work is part of the program of the ESRC Centre for Economic Learning and Social Evolution, University College London.

Correspondence concerning this article should be addressed to Francisco J. López, Departamento de Psicología Básica, Psicobiología y Metodología, Facultad de Psicología, Campus de los Teatinos, Universidad de Málaga, Málaga 29071, Spain. Electronic mail may be sent to fj_lopez@sci.uma.es.

rently, participants received $DE \rightarrow O_2$, $E \rightarrow O_2$, and $F \rightarrow$ no O trials. After a number of trials, participants estimated the relationship between cue A and outcome O_1 and between cue D and outcome O_2. The probability of outcome O_1 given cue A, $P(O_1/A)$, is 1.0 (because disease 1 always occurs on AB trials), and $P(O_1/\overline{A})$ is .2 (because disease 1 occurs in 1 out of the 5 trial types in which cue A is absent, and these trial types are equiprobable). According to the ΔP rule, then, the contingency between cue A and outcome 1 was 0.8. In terms of ΔP the contingency between D and O_2 was also .8 [$P(O_2/D) = 1.0$ and $P(O_2/\overline{D}) = .2$], yet people gave reliably greater judgments for the $A \rightarrow O_1$ than for the $D \rightarrow O_2$ relationship, and common sense suggests that this result is reasonable: O_1 is contingent on cue A because it does not occur in the presence of cue B alone, whereas O_2 is not contingent on cue D because it also occurs on trials with E alone.

This finding is at variance with the ΔP rule, but as Cheng and Holyoak (1995) have pointed out, this measure, as computed in Equation 1, is not the appropriate normative standard against which individuals' judgments should be evaluated. In fact, a reinterpretation of this rule allows these particular results and related cue selection effects to be explained (Cheng & Holyoak, 1995; Melz, Cheng, Holyoak, & Waldmann, 1993). This new version of the normative theory is called the probabilistic contrast model (Cheng & Novick, 1990, 1991, 1992).

Probabilistic Contrast Model

The probabilistic contrast model (PCM) represents a normative extension of the ΔP metric to situations in which the target cue–outcome relationship is evaluated against a varying context (i.e., situations where multiple predictors co-occur or the background cues change). In these more complex situations, both the relationship between a target cue (A) and the outcome and also the predictive status of other cues are taken into account. Specifically, the contingency between cue A and the outcome must be computed over certain restricted *focal sets* of events. Suppose that cue A co-occurs with cue B and both cues are paired with the outcome. To evaluate the true or normative relationship between cue A and the outcome, the status of any other possible alternative predictors of the outcome (such as B) need to be held constant. This requirement is met if the contingency between A and the outcome is computed over a focal set in which cue B is always present or over a focal set in which cue B is always absent. Then, the contingency between A and the outcome conditional on the presence of the alternative cue B, ΔP_B, is simply the difference between the proportion of cases in which the outcome is present given the presence of both cues and the proportion of cases in which the outcome is present given the absence of the target cue A and the presence of the alternative cue B:

$$\Delta P_B = P(O/A \cdot B) - P(O/\overline{A} \cdot B). \quad (2)$$

Alternatively, the contingency between A and the outcome may be computed as the difference between the proportion of cases in which the outcome is present given the presence of the target cue A and the absence of the alternative cue B and the proportion of cases in which the outcome is present given the absence of both cues:

$$\Delta P_{\overline{B}} = P(O/A \cdot \overline{B}) - P(O/\overline{A} \cdot \overline{B}). \quad (3)$$

Thus, the PCM comes to satisfy a normative requirement about when two events may be regarded as related. The evaluation of the predictive status of a cue must be based on a contrast between what happens when it is present versus what happens when it is absent, all else being held constant. Cheng and Holyoak (1995) have recently put forward a specific implementation of this normative analysis to explain how people detect predictive relationships between events. It should be noted that the PCM was originally presented as a computational-level theory (e.g., Cheng & Novick, 1992) with no specification of how the computations might be performed in mental processing terms. In the present article, we were principally concerned with the algorithmic-level version of the PCM formulated by Cheng and Holyoak and Melz et al. (1993) and directly contrasted by them with the Rescorla–Wagner theory. We refer to this as the PCM-ALG model. Cheng and Holyoak's implementation of the PCM includes several steps: (a) selecting which cues will be incorporated in the focal sets or will be used to conditionalize other cues, (b) choosing the conditional contrasts to be calculated, and (c) integrating the information provided by the different conditional contrasts into a single judgment about the predictive value of the target cue. According to this particular version, the different focal sets are exclusively defined in terms of the normative requirements mentioned above.

To see how the PCM-ALG implementation works, let us consider its application to the experiment by Shanks (1991, Experiment 2). With regard to the target cue A–outcome O_1 relationship, the model implementation predicts that cue C will be initially selected as a conditionalizing cue given that it has been consistently paired with outcome O_1. Though cue B might in principle be regarded as a potential conditionalizing cue, the fact that cue B was consistently paired on its own with the absence of the outcome precludes its consideration as a potential alternative predictor. As participants have no prior beliefs about the predictive value of any other cue, cue C may be initially regarded as the only conditionalizing cue. Hence, this is the cue whose status (presence or absence) needs to be controlled to correctly assess the true predictive value of the target cue A. In general, people will prefer the contrast conditional on the absence of all conditionalizing cues (for a presumably positive relationship between the target cue and the outcome) or the contrast conditional on the presence of an already known predictive cue (for a presumably negative relationship; see Cheng & Holyoak, 1995, for further explanation of this point). From the information provided in the task, the only meaningful conditional contrast that can be computed is the one that is conditional on the absence of cue C. According to this contrast, participants will compute the probability of O_1 given the presence of A and the absence of C, $P(O_1/A \cdot \overline{C})$,

which has a value of 1.0. In addition, participants will compute the probability of O_1 in the absence of both A and C, $P(O_1/\overline{A} \cdot \overline{C})$, which has a value of 0. Thus, cue A has a conditional contingency of 1.0 in the focal set in which cue C is absent.

With regard to the relationship between the target cue D and outcome O_2, cue E will be selected as the only conditionalizing cue as it has been consistently paired with outcome O_2. Again, from the information provided in the task, the only conditional contrast that can be computed is the one conditional on the presence of cue E. Thus, participants will compute the probability of O_2 given the presence of both D and E, $P(O_2/D \cdot E)$, which has a value of 1.0. In addition, they will compute the probability of O_2 in the absence of cue D and the presence of cue E, $P(O_2/\overline{D} \cdot E)$, which also has a value of 1.0. Therefore, cue D has a conditional contingency of 0 in the focal set in which cue E is present. Because the most relevant conditional contrast, the contrast conditional on the absence of cue E, cannot be computed, participants will not be completely certain about the predictive value of cue D. In this case, they may partially consider the unconditional ΔP metric between cue D and outcome O_2, which yields a value of 0.8. However, it should be remembered that this is tantamount to saying that participants consider an inappropriate or nonnormative contingency measure for the situation involved here. In accordance with Cheng and Holyoak's (1995) implementation, participants will integrate the information from the different contrasts that have been computed. Thus, ratings for cue D will vary between 0 and .8 (on a scale from 0 to 1.0). Hence, higher ratings are expected for cue A than for D, which was the empirical result observed.

The Associative Account

Another way of explaining these results is based on an entirely different mechanism. According to associative models, people form mental connections between the various events and update these connections on a trial-by-trial basis according to some connectionist learning rule. One simple rule was proposed by Rescorla and Wagner (1972) to deal with associative learning in conditioning experiments, and is formally equivalent to the delta rule used in many connectionist learning models (Gluck & Bower, 1988; Kruschke, 1992, 1993; McClelland & Rumelhart, 1985). On each trial, the strength (V) of the target association is changed by an amount ΔV, given by:

$$\Delta V = \alpha \times \beta \times (\lambda - \Sigma V), \quad (4)$$

where α and β are learning rate parameters determined by the saliences of the cue ($\alpha = 0$ when the cue does not occur on a trial) and the outcome, respectively; λ is the total association strength that can be supported by the outcome (normally set to 1.0 if the outcome occurs on a given trial and to 0.0 otherwise); and ΣV is the sum of the associative strengths of all cues present on the current trial. Learning continues until λ equals ΣV, which means that the outcome is fully predicted on trials on which it occurs and its absence is also fully predicted on trials on which it does not occur. In those cases in which the cue predicts the absence of the outcome, its associative strength becomes negative. It is assumed that people's ratings of the relationship between the target cue and the outcome are monotonically related to the associative strength of that cue (V).

How does the Rescorla–Wagner theory explain the results of the Shanks (1991, Experiment 2) study? In the contingent condition, the model assumes that the associative strength of both cues A and B will increase on $AB \rightarrow O_1$ trials because they are paired with outcome O_1. However, the associative strength of cue B will decrease on $B \rightarrow$ no O trials as this cue is explicitly paired with the absence of the outcome. Eventually, the competitive nature of the learning algorithm will ensure that cue A ends up with an associative strength of λ. In the noncontingent condition, the associative strengths of both cues D and E will increase on $DE \rightarrow O_2$ trials. In addition, cue E will further augment its associative strength on $E \rightarrow O_2$ trials. Note that when ΣV begins to exceed λ on $DE \rightarrow O_2$ trials, both cues will lose associative strength. Yet, cue E will still increase its strength on $E \rightarrow O_2$ trials. As a consequence of this trade-off, cue E will end up with all the associative strength available (λ) and the associative strength of cue D will be zero. Trials with cues C and F have been dropped from this analysis because they do not affect what is learned about the other cues. Hence, the theory predicts a greater associative strength for cue A in the contingent condition than for cue D in the noncontingent one. Thus, higher ratings are expected for cue A than for D, consistent with people's ratings.

Comparison of the Theories

To summarize the discussion to this point, two accounts of how contingency judgments are made have been described, one based on the computation of conditional contrasts and the other based on accumulating changes in associative strength. How can we decide between these theories? A large amount of research has been conducted with the aim of distinguishing them (e.g., Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Chapman, 1991; Cheng & Holyoak, 1995; Melz et al., 1993; Price & Yates, 1993; Shanks, 1991, 1993a; Van Hamme & Wasserman, 1994; Wasserman et al., 1993; Waldmann & Holyoak, 1992; see Allan, 1993, and Shanks, 1993b, 1995a, for reviews). In what follows, some of these efforts are briefly reviewed.

One strategy has been to obtain from participants not only their contingency judgments but also their ratings of the relevant conditional probabilities, and then to evaluate the extent to which the function relating them is as described by Equation 1. In an instrumental learning task, Wasserman et al. (1993) asked participants to rate an action–outcome relationship and at the same time to estimate the probability of the outcome (O) in the presence [$P(O/A)$] and in the absence [$P(O/\overline{A})$] of the action (A). In different problems the actual value of ΔP varied from -1.0 to 1.0. In these experiments the target action took place against a nonvarying context; hence, the unconditional ΔP measure is the

appropriate normative metric of the programmed relationships.

Although contingency judgments were generally accurate, and although probability estimates were generally quite close to the programmed probabilities, Wasserman et al. (1993) concluded that contingency judgments were not mediated by estimates of the relevant conditional probabilities, but were instead better accounted for by the Rescorla–Wagner theory. The reason for reaching this conclusion was that biases in contingency estimates were not well explained by biases in probability estimates. For instance, on the basis of participants' probability estimates, increasing $P(O/A)$ from 0.0 to 1.0 should have had a smaller impact on contingency judgments than increasing $P(O/\overline{A})$ from 0.0 to 1.0. In fact, the exact opposite effect was observed in individuals' contingency ratings.

Another piece of evidence that can be marshaled in favor of the associative theory comes from studies of the learning functions obtained by asking people to estimate event contingencies after varying numbers of trials. In several studies, a relationship was established between a target cue and an outcome within a constant context, and the general finding (e.g., Shanks, 1985a, 1987, 1995a; Shanks, López, Darby, & Dickinson, 1996) is that under a positive contingency, judgments start close to zero and slowly increase in a negatively accelerated fashion toward the actual contingency, whereas under a negative contingency, judgments slowly decrease across trials toward the actual contingency. When ΔP is zero but the probability of the outcome is fairly high, an interesting finding is that judgments tend to increase initially before falling back toward zero (e.g., Shanks et al., 1996), a result that is also obtained in animal conditioning studies (Kremer, 1971; Rescorla, 1972). The overall pattern of results is well accommodated by the Rescorla–Wagner theory, because associative strengths are assumed to start at zero and because increments and decrements in associative strength get smaller the closer ΣV is to its asymptotic level (λ). The preasymptotic biases found in participants' ratings when ΔP is zero and outcomes are very frequent is also predicted by the Rescorla–Wagner theory (see Shanks, 1995a, for further explanation). Although learning functions are not necessarily incompatible with a normative account (e.g., Fales & Wasserman, 1992, have suggested that Bayes's theorem may accommodate such functions), these results are problematic for theories that assume that contingency judgments are based on ΔP, because the latter was kept constant across trials in these experiments. In addition, it is not obvious how a normative analysis can give an account of the preasymptotic bias shown by judgments in high-frequency noncontingent conditions.

Although these results tend to favor the Rescorla–Wagner theory, some data reported by Waldmann and Holyoak (1992) seem to be better explained by the PCM. Waldmann and Holyoak have supplemented the contingency account with a causal model theory. According to this theory, people are sensitive to the causal status of cues and outcomes when learning predictive relationships between events. Specifically, this theory predicts that in a situation in which multiple cues are possibly related to a single outcome (as in the previously mentioned experiment by Shanks, 1991), cue selection effects may only emerge when the cues are interpreted as causes and the outcome as an effect (e.g., different foods [cues] causing an allergic reaction [outcome]) but not when the cues and the outcome are interpreted as effects and a cause, respectively (e.g., different symptoms [cues] being caused by a virus [outcome]). This comes about because, as Waldmann and Holyoak (1992) said, "People have a strong predisposition to learn directed links from causes to effects, rather than vice-versa, even in situations in which they receive effect information prior to cause information" (p. 224) and "different effects, like different dependent measures obtained in an experiment, do not compete with one another; rather, each effect, as well as any interaction among the effects, provides information about the consequences of the cause" (p. 226). Because associative theories ignore the possible causal status of events and merely focus on their relatedness, it should make no difference whether a cue–outcome learning task is presented in an effect → cause rather than a cause → effect format. Waldmann and Holyoak obtained evidence favoring their causal model theory over associationist theories. However, their study has been criticized on methodological grounds and, contrary to the causal model theory, evidence has been reported that cue competition does emerge in effect → cause situations (Matute, Arcediano, & Miller, 1996; Price & Yates, 1995; Shanks & López, 1996; see also Waldmann & Holyoak, 1997). Thus, the evidence on this issue is currently inconclusive.

Two key points are illustrated by this short review. First, the two theories make divergent and testable predictions in a wide range of situations. Second, although previous research has not succeeded in yielding a conclusive answer to the issue of which theory better accounts for human contingency judgment data, the associative explanation seems to be favored, in general, by the current data. Nevertheless, we believe that firmer conclusions may be reached from studies focusing on different aspects of contingency judgment, specifically, those aspects related to the nature of the mechanism by which people compute predictive relationships.

Trial Order Effects

Perhaps one of the clearest ways of distinguishing between the two theories is in terms of their predictions concerning manipulations of the order in which different trial types are seen. If people compute event relationships by maintaining accurate records of the normatively relevant conditional probabilities, changing the order in which trials are seen should, according to the PCM-ALG, have no effect on contingency estimates: Such conditional probabilities are the same regardless of the order in which the events are observed. The Rescorla–Wagner theory, on the other hand, predicts profound effects of trial order variations. The predictions of the model are elaborated in detail later, but for present purposes the main thing to note is that the model predicts *recency–density* effects whereby the most recent and dense trial types will have the biggest impact on judgments.

Trial order effects are inherently interesting because they allow a direct examination of the processes involved in the real-time updating of beliefs. Although trial order effects have been demonstrated in a variety of judgment tasks (see Hogarth & Einhorn, 1992), the evidence for order effects in contingency judgment is very sparse. Some of the best evidence comes from a series of studies by Chapman (1991), which focused on the selectional effect called *blocking* (Kamin, 1968). In her experiments, which as in Shanks's (1991) study used a medical diagnosis procedure, it was shown that contingency judgments were affected by trial order manipulations. For instance, in Chapman's Experiment 2, participants received trials divided into three phases. Trials in Phase 1 showed that cue P predicted outcome O_1 ($P \rightarrow O_1$) and cue N was paired with the absence of the outcome ($N \rightarrow$ no O). In Phase 2, the combination of cues P and F predicted O_1 ($PF \rightarrow O_1$) and the combination of cues N and B also predicted O_1 ($NB \rightarrow O_1$). Then in Phase 3, N was paired with outcome O_1 ($N \rightarrow O_1$) and cue P was paired with the absence of the outcome ($P \rightarrow$ no O). Note that PF and NB were treated identically in Phase 2, and that both P and N were associated by themselves with O_1 in one phase and were not associated with O_1 in another phase; for P the association with O_1 occurred in Phase 1, whereas it occurred in Phase 3 for N. Because the relevant trial types for the target $F \rightarrow O_1$ and $B \rightarrow O_1$ relationships are, therefore, identical, except that the order in which these trials are presented is different, the PCM-ALG predicts that ratings for these two cues should be equal. However, participants rated the predictive value of cue B as being greater than that of cue F. Although inconsistent with contingency-based theories, this result is predicted by the Rescorla–Wagner theory. According to that theory, symptom P will have acquired positive associative strength in the first stage and will therefore block symptom F from acquiring significant associative strength on the $PF \rightarrow O_1$ trials. In contrast, symptom B will acquire half of the associative strength available on the $NB \rightarrow O_1$ trials, and this will not be affected by the subsequent $N \rightarrow O_1$ trials.

Although this particular result only contradicts the PCM-ALG's predictions, other trial order effects reported by Chapman (1991) were problematic for both the PCM-ALG and the Rescorla–Wagner theory. For example, in her Experiment 3, Chapman found that blocking could be demonstrated in situations where compound cue trials are presented before (rather than after) single cue type trials, a result that is not predicted by the Rescorla–Wagner theory. At variance with the PCM-ALG, this procedure generated a blocking effect of a significantly lesser magnitude than the standard procedure, however. Thus, individuals re-evaluated the predictive values of the target cues as a result of new information about the predictive value of competing cues (see also Shanks, 1985b, for similar results).

Probably the main question in studies of trial order effects in the updating of beliefs is whether judgments are most influenced by information presented first or last (i.e., primacy vs. recency). Chapman's (1991) findings are hard to interpret in terms of primacy and recency because trial order for the target cues was not directly manipulated. In Chapman's Experiment 2, for example, the order of trials paired with the target cues F and B was manipulated, but not that of F and B themselves. Only a small number of studies have asked whether primacy or recency effects occur in contingency ratings. In one study (Yates & Curley, 1986; see also Curley, Young, Kingry, & Yates, 1988), participants had to detect the relationship between the color of a fictitious plant and its region of origin. Although the overall information provided indicated that these attributes were unrelated (i.e., $\Delta P = 0$), trials were divided into two blocks containing contradicting information about this relationship: In one condition, the early information suggested a positive relationship and the later information a negative one, whereas in a second condition, this ordering was reversed. In addition to having to rate the predictive relationship between the attributes, participants had to explicitly rate the conditional probabilities that defined the contingency (see Equation 1). The results showed that contingency judgments derived from these latter conditional probability ratings showed a primacy effect, but this effect only occurred when participants were not informed at the beginning of the experiment about the later rating test. Participants informed about the test showed no order effect. Wasserman, Kao, Van Hamme, Katagiri, and Young (1996) also failed to observe order effects in two experiments in which participants judged the relationship between the presence or absence of a fertilizer and whether or not a plant bloomed. As in the Yates and Curley (1986) study, ΔP was zero overall but went from positive to negative in some conditions and from negative to positive in others.

In general, these results show that people's contingency judgments are often insensitive to the order in which trials are presented, as the PCM-ALG predicts. However, small effects are sometimes observed, as in Chapman's (1991) experiments. The experiments by Yates and Curley (1986) and Wasserman et al. (1996) are probably not very diagnostic because quite small numbers of trials were used and thus there is little opportunity for the effects of a belief updating mechanism to become apparent. Given the theoretical importance of these effects in potentially distinguishing between contingency-based and associative accounts of contingency judgment, further explorations of their importance on such judgments is in order. It is perhaps also worth mentioning that in situations where order effects are commonly observed, for example in studies of personality judgment (Hogarth & Einhorn, 1992), they have in general proven difficult to account for. Judgment models based on Bayes's theorem, for instance, predict no order effects under any circumstances (Slovic & Lichtenstein, 1971).

Overview of the Experiments

The four experiments we present specifically evaluated the possible influence of trial order on contingency judgments using longer series of trials. In all of the experiments, individuals performed a trial-by-trial learning task in which they had to learn predictive relationships between different cues and outcomes. Subsequently, in a test stage, they had to rate some target cue–outcome relationships. In the first three

experiments, during the learning stage, trials were divided into two distinct blocks. The relationships programmed in these two blocks were different: In fact, they contradicted each other, though to different extents in the different experiments. Thus, two trial order conditions were defined according to the particular block of trials that was presented first. At the end of the learning stage, taken as a whole, the statistical relationships arranged between the different cues and outcomes were identical for both trial order conditions. Hence, any difference in participants' contingency judgments would reflect the impact of trial order on these ratings, as this was the only difference between the experimental conditions. One alternative possibility is that participants simply forgot the relationships programmed during the first block of trials and, thus, different ratings are expected in the two trial order conditions. This possibility was specifically evaluated in Experiment 3. Another possibility relates to the fact that trials were presented in two distinct blocks and, hence, any difference between judgments might be due to this aspect of the procedure. In Experiment 4, therefore, a gradual transition was made between blocks in the sense that trial types were intermixed in a particular way, though two trial order conditions were still defined. In addition, this design allowed us to evaluate whether participants were operating on a limited moving-window basis, so that only the most recent information was considered in their contingency ratings. Both of these possibilities can be regarded as alternative interpretations of trial order effects that are compatible with the PCM-ALG's predictions, provided the model is supplemented with additional mechanisms.

Experiment 1

In Experiment 1 we evaluated whether the order in which trials are presented influences contingency judgments. The task was divided into two blocks of trial types, and two trial order conditions were compared. In one of the trial order conditions, a contingent relationship was programmed between a target cue and an outcome during the first block of trials, whereas in the second block this relationship was turned into a noncontingent one. In the other trial order condition, the order in which blocks were presented was simply reversed.

As in Shanks' (1991) experiment, participants received information about the symptoms that a particular patient presented and they had to diagnose the disease this patient was suffering from. All of the relationships were probabilistic. Table 1 shows that the cues of interest, A and D, were followed by their respective outcomes O_1 and O_2 exactly the same number of times and in compound with cues that had undergone exactly the same treatment. The only difference between A and D was that A was a better predictor of O_1 than its pairmate B in Block 1 and a worse predictor than its pairmate C in Block 2, whereas for D this was reversed. That is, D was a worse predictor of O_2 than E in Block 1 and a better predictor than F in Block 2. After seeing all of the trials, participants judged the relationship between cue A and outcome O_1 and between cue D and outcome O_2. If these two ratings differed, judgments could be regarded as being sensitive to trial order, as this is the only factor in which the two conditions differed.

The contingency-based and the associative model make different predictions about contingency ratings in the two conditions. Specifically, the PCM-ALG predicts that judgments should be identical across conditions, whereas the Rescorla–Wagner model predicts lower judgments in the contingent–noncontingent (con–non) condition than in the noncontingent–contingent (non–con) condition. Where do these model predictions come from? With regard to the PCM-ALG, judgments concerning the $A \rightarrow O_1$ and $D \rightarrow O_2$ relationships should not differ because the probabilistic evidence provided during the task in these two conditions is identical (see the Appendix for detailed contingency calculations as predicted by the PCM-ALG). In contrast, the Rescorla–Wagner theory predicts a recency effect in individuals' ratings. The dotted lines of Figure 1 show the associative strengths predicted for target cues A and D across trials under the two experimental conditions (the meaning of the continuous lines is discussed below). With respect to the $A \rightarrow O_1$ relationship (con–non condition), the associative strength of cue A increases during the first block of trials until reaching an asymptotic positive value of about .6. During the second block of trials, its associative strength diminishes towards an asymptotic value of 0, provided a sufficient number of trials is presented. With regard to the $D \rightarrow O_2$ relationship (non–con condition), the associative strength of cue D at the end of the first block of trials is 0. However, its associative strength increases until reaching an asymptotic value of around 0.6 during the second block of trials. Hence, lower ratings are expected for cue A than for cue D at the end of the learning phase, as participants'

Table 1
Design of Experiment 1

Trial order condition	Block 1	Block 2	Test cue	Judgment
Con–non	8 $AB \rightarrow O_1$/2 $AB \rightarrow$ no O 8 $B \rightarrow$ no O/2 $B \rightarrow O_1$	8 $AC \rightarrow O_1$/2 $AC \rightarrow$ no O 8 $C \rightarrow O_1$/2 $C \rightarrow$ no O	$A \rightarrow O_1$?	$M = 73.6$
Non–con	8 $DE \rightarrow O_2$/2 $DE \rightarrow$ no O 8 $E \rightarrow O_2$/2 $E \rightarrow$ no O	8 $DF \rightarrow O_2$/2 $DF \rightarrow$ no O 8 $F \rightarrow$ no O/2 $F \rightarrow O_2$	$D \rightarrow O_2$?	$M = 80.6$

Note. Con = contingent; Non = noncontingent. A, B, C, D, E, and F are symptoms; O_1 and O_2 are diseases 1 and 2, respectively; no O = no disease.

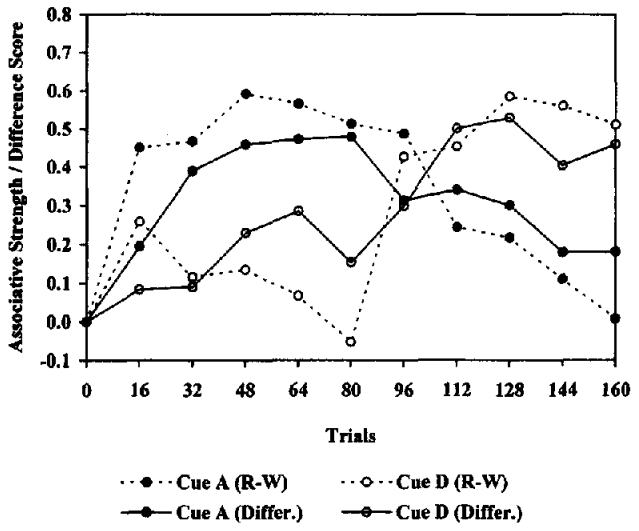

Figure 1. Solid lines: differences (Differ.; cue *A*) across the learning stage between the proportion of times disease 1 is predicted on *AB* versus *B* and *AC* versus *C* trials and differences (cue *D*) between the proportion of times disease 2 is predicted on *DE* versus *E* and *DF* versus *F* trials in Experiment 1. Dashed lines: predicted changes in the associative strengths of cues *A* and *D* across the learning stage according to the Rescorla–Wagner (R-W) theory. Predictions were averaged across 10 replications of the same number and equivalent pattern of training trials as the participants received in Experiment 1. All α and β parameters were set to .2, and λ was set to 1.0 and 0 for the occurrence and absence of the target outcome, respectively.

ratings about the predictive value of a particular cue are monotonically related to its associative strength.

Method

Participants and apparatus. The participants were 36 psychology undergraduates from Universidad de Málaga who volunteered to take part in the experiment. All participants were tested individually, and the task was presented on an IBM-PC computer connected to a color monitor. Participants' judgments were requested by means of a paper-and-pencil questionnaire.

Procedure. After a brief verbal description of the task, participants read the instructions on the computer screen. Their task was to diagnose a set of hypothetical patients according to the symptoms they had. A total of 160 trials was presented, divided into two blocks of 80 trials. Each trial comprised the case history of a single patient. Table 1 displays the trial types that were programmed, and these were randomized within blocks. A total of 10 trials of each trial type was presented. The two blocks were presented in a single series with no break so that participants could not formally distinguish them.

Table 1 displays only half of the trial types that occurred in the experiment. The other half were identical except that different symptoms (*G* to *L*) and diseases (3 and 4) were used. The relationships between these new symptoms and diseases mimicked the ones described in Table 1. Thus, there were two sets of relationships for each of the con–non and non–con conditions. For half of the participants, symptoms *A–L* were swollen gums, coughing, painful breathing, fever, excessive perspiration, sickness, headache, tremor, baldness, tachycardia, sight loss, and itchiness, respectively. For the remaining participants, the assignments for symptoms *A* and *D* were reversed and also for symptoms *G* and *J*. Orthogonally to this, for half of the participants, diseases 1–4 were Cajal's Disease, Hocitosis, Ochoa's Syndrome, and Beralgia, respectively. For the other half, the assignments for diseases 1 and 2 were reversed and also for diseases 3 and 4.[1] On each trial, participants read on the screen the symptom or symptoms that the patient had: "Patient number 'X' has the following symptoms. . . ." Then, a list of symptoms appeared and the correct diagnosis for that patient was requested: "What is the correct diagnosis?" There were five possible diagnostic categories for each patient: M = Cajal's Disease, H = Hocitosis; S = Ochoa's Syndrome, B = Beralgia, and N = no disease. After participants had made their diagnosis, the program indicated what the correct diagnosis was for that particular patient: "The correct answer is Cajal's Disease/Hocitosis/Ochoa's Syndrome/Beralgia/no disease." A beep accompanied the correct diagnosis if the participant's diagnosis was wrong. For symptoms *AB*, the correct diagnosis was disease 1 on 80% of the trials and no disease on the remaining 20%. For symptoms *DE*, the correct diagnosis was disease 2 on 80% of the trials and no disease on the remaining 20%, and so on (see Table 1). Participants pressed the space bar to see the next patient.

After all of the trials had been presented, participants had to rate on a scale from 0 to 100 the relationships between the target symptoms and each of the diseases: "Mark with an 'X' on a scale from 0 to 100 the degree of relationship between [symptoms *A/D*] and the different diseases. Zero means that there is no relationship and 100 means a perfect relationship." Although participants rated the degree of relationship between each symptom and disease, only those judgments concerning the target relationships are considered, namely symptom $A \rightarrow$ disease 1 and symptom $D \rightarrow$ disease 2. As two identical sets of symptoms and diseases were included per trial order condition and a counterbalancing procedure was adopted, participants' judgments were collapsed into a single measure, one per trial order condition.

Results and Discussion

Table 2 shows the percentage of correct diagnoses across participants for the last two trials of each trial type at the end of each training block. According to these mean percentages (M = 55.2 and M = 59.6 for Blocks 1 and 2, respectively), the task did not raise special difficulties for the participants. As there were five response categories, chance performance is 1 out of 5, or 20% correct. Given the probabilistic nature of the task, the maximum level of correct responding achievable is 80%.

All the statistical analyses reported in this series of experiments adopted a significance level of 99% (α = .01), unless otherwise stated. The critical data are the judgments made at the end of the learning stage. Table 1 shows that ratings were higher in the non–con than in the con–non condition (M = 80.6, SE = 3.7 and M = 73.6, SE = 4.4, respectively). A t test showed that the difference was

[1] All experiments were conducted in Spanish. The actual Spanish names for symptoms *A–L* were *encías inflamadas, tos, respiración fatigosa, fiebre, sudoración excesiva, mareos, dolor de cabeza, temblor, caída del cabello, taquicardia, pérdida de visión,* and *picor,* respectively. The Spanish names for diseases 1–4 were *mal de Cajal, hocitosis, síndrome de Ochoa,* and *beralgia,* respectively.

Table 2
Percentages of Correct Diagnoses on the Last Two Occurrences of Each Trial Type in the Learning Stage of Experiment 1

Trial type	% correct
Block 1	
AB	55.6
B	58.3
DE	59.7
E	47.2
M	55.2
Block 2	
AC	67.4
C	50.7
DF	63.4
F	56.9
M	59.6

Note. Chance performance is 20%. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms.

statistically significant, $t(35) = 3.11$. Thus, the results showed a significant trial order effect on contingency judgments. Contrary to the predictions of the PCM-ALG, and in contrast to the findings of Yates and Curley (1986) and Wasserman et al. (1996), participants' judgments were affected by the order in which trials were presented and showed a recency effect. In addition, the specific trial order effect obtained is predicted by the Rescorla–Wagner model, as shown in Figure 1 (dotted lines).

As can be seen in these dotted lines of Figure 1, the Rescorla–Wagner theory postulates gradual changes in the associative strength of the target cues across trials. The model predicts that ratings at the end of Block 1 should be higher in the con–non than in the non–con condition, and this should be followed by an inversion of the relative ordering of ratings for the two trial order conditions by the end of the task. Although judgments were only elicited at the end of the learning stage so as not to make the trial blocks distinguishable, an indirect measure of how individuals' estimations of the target relationships evolve across trials can be derived. Thus, the Rescorla–Wagner theory's predictions across trials can also be tested. The model assumes that the difference between the proportion of times disease 1 is diagnosed on *AB* trials and on *B* trials reflects the associative strength of cue *A* across the first block of trials. Similarly, the difference between the proportion of times disease 1 is diagnosed on *AC* trials and on *C* trials reflects the associative strength of cue *A* during the second block of trials. Correspondingly, the difference between the proportion of times disease 2 is diagnosed on *DE* trials and on *E* trials indexes the associative strength of cue *D* across the first block and that between *DF* and *F* trials its associative strength across the second block. Hence, we can indirectly track through these differences how participants' estimations of the target relationships evolved during the learning session. The continuous lines of Figure 1 show that these difference scores mirror quite closely the predicted associative strengths.

A 2 (order: con–non vs. non–con) × 10 (blocks) within-subjects analysis of variance (ANOVA) was performed on these difference scores. The blocks variable included two trial presentations of each target trial type. Only blocks and the Order × Blocks interaction produced significant differences, $F(9, 315) = 5.67$, $MSE = 0.09$, and $F(9, 315) = 9.50$, $MSE = 0.10$, respectively. Four nonorthogonal planned contrasts served to test the associative model's most relevant predictions. In the con–non condition, the difference score after the fifth block was significantly greater than after the first block, $F(1, 35) = 15.09$, $MSE = 0.10$. In addition, the difference score after the last block was smaller than after the fifth block, $F(1, 35) = 14.08$, $MSE = 0.10$. In the non–con condition, the difference score between the fifth and first block was not significant, $F(1, 35) = 0.70$, $MSE = 0.10$, but the difference score after the last block was greater than after the fifth block, $F(1, 35) = 14.91$, $MSE = 0.10$. As indicated by these contrasts, the main features of participants' differential diagnoses across trials can be readily understood in terms of the associative model.

Experiment 2

The results from Experiment 1 show the influence of trial order on contingency judgments, contrary to the predictions of the PCM-ALG. The PCM-ALG assumes that apart from the selection of restricted focal sets, all trial types contribute to the computation of contingency. The order in which these trials is presented should have no effect.

Nevertheless, ratings were very high in both trial order conditions. In fact, mean judgments were above 70 on a scale from 0 to 100, and the actual trial order effect was small in absolute terms. This result is in line with the results of previous research, as mentioned above, in which empirically small trial order effects were also found (Chapman, 1991; Shanks, 1985b).

In our previous experiment, cues went from being reliable predictors of the outcomes to being nonpredictors of these outcomes, or vice versa. Although the changes in the relationships programmed between blocks were large enough to produce a significant order effect, more dramatic changes in these relationships should create a greater effect. Such modifications in the design do not qualitatively alter the main predictions made by the PCM-ALG or Rescorla–Wagner theory. Hence, evaluating the influence of trial order in this new situation should be more illuminating for comparing the predictions of the two theories.

Table 3 shows the design of the present experiment in which a more radical manipulation of trial order is made. As before, two trial order conditions were programmed. Cue *A*'s contingency for outcome O_1 went from strongly positive in Block 1 to strongly negative in Block 2, whereas for cue *D* these contingencies (for outcome O_2) were reversed. As with Experiment 1, the PCM-ALG predicts no order effect on contingency judgments because the trial types for the two conditions are identical (contingency calculations as predicted by PCM-ALG are also included in the Appendix). In contrast, the Rescorla–Wagner theory again predicts a strong recency effect in contingency judgments. The dotted lines of Figure 2 present the associative strengths predicted for the

Table 3
Design of Experiments 2 and 3

Trial order condition	Block 1	Block 2	Test cue	Judgments Exp. 2	Judgments Exp. 3
Con–inh	$AB \rightarrow O_1$ $B \rightarrow$ no$O$	$AC \rightarrow$ no$O$ $C \rightarrow O_1$	$A \rightarrow O_1$?	$M = 19.7$	$M = 18.2$
Inh–con	$DE \rightarrow$ no$O$ $E \rightarrow O_2$	$DF \rightarrow O_2$ $F \rightarrow$ no$O$	$D \rightarrow O_2$?	$M = 69.4$	$M = 45.0$

Note. Exp. = experiment; con = contingent; inh = inhibitory. A, B, C, D, E, and F are symptoms; O_1 and O_2 are diseases 1 and 2, respectively; no O = no disease.

target cues A and D across trials under the two experimental conditions. As can be seen, the theory predicts that judgments in the contingent–inhibitory (con–inh) condition should be lower than those in the inhibitory–contingent (inh–con) condition. In addition, this associative account is sensitive to the fact that a more radical change in the contingencies programmed within each condition has been made and, hence, a greater trial order effect is expected for this design. Specifically, concerning the $A \rightarrow O_1$ relationship, the associative strength of cue A increases during the first block of trials toward an asymptotic positive value of 1.0. During the second block of trials, its associative strength diminishes toward an asymptotic negative value of -1.0. This comes about because on $AC \rightarrow$ no O trials, $\lambda = 0$ and ΣV is greater than 0 as a result of the $C \rightarrow O_1$ trials and, therefore, cue A's associative strength will acquire a negative value, assuming a sufficient number of trials is provided.

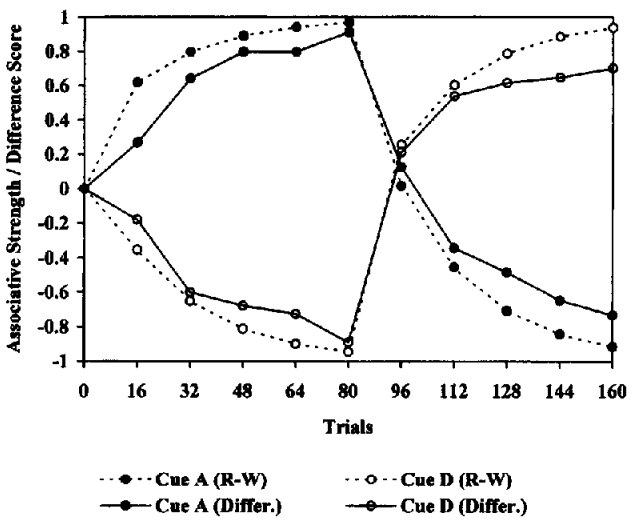

Figure 2. Solid lines: differences (Differ.; cue A) across the learning stage between the proportion of times disease 1 is predicted on AB versus B and AC versus C trials and differences (cue D) between the proportion of times disease 2 is predicted on DE versus E and DF versus F trials in Experiment 2. Dashed lines: predicted changes in the associative strengths of cues A and D across the learning stage according to the Rescorla–Wagner (R-W) theory. Predictions were computed using the procedure described in Experiment 1 (see Figure 1).

With regard to the $D \rightarrow O_2$ relationship, the asymptotic associative strength of cue D at the end of the first block of trials is -1.0, as it reliably predicts the absence of O_1. However, its associative strength increases toward an asymptotic value of 1.0 during the second block of trials. Hence, lower ratings are expected for cue A than for cue D. Given that the difference between the associative strengths of the target cues at the end of the learning stage is larger now than in Experiment 1, the experimental manipulation should have a greater impact on participants' judgments in Experiment 2.

Method

Participants and apparatus. The participants were 32 psychology undergraduates from Universidad de Málaga who volunteered to take part in this experiment. They had not participated in Experiment 1. The same computer and similar questionnaires were used as in Experiment 1.

Procedure. Except where specifically mentioned, procedural details were as for Experiment 1. The experiment was again presented as a medical diagnosis task.

Table 3 shows half of the trial types that occurred during the experiment. The other half was identical except that different symptoms (G to L) and diseases (3 and 4) were used. The symptom and disease names as well as the counterbalancing procedure were as described for Experiment 1.

At the end of the learning phase, participants rated the relationships perceived between the target symptoms and each of the diseases on a scale from -100 to 100. The negative half of the scale was introduced because of the negative relationships programmed between cues and outcomes. Specific written instructions were provided about the exact meaning of this negative scale:

> Use the POSITIVE HALF of the scale if you think that the symptom indicates the presence of the disease. Use the NEGATIVE HALF if you think that the symptom indicates the absence of the disease. Use values close to 0 if you think that symptoms and diseases are unrelated.

Results and Discussion

Table 4 shows the percentage of correct diagnoses across participants for the last two trials of each trial type at the end of each block. By the end of each block, participants had learned reasonably well the relationships programmed in that block ($M = 93.4$, $SE = 1.5$ and $M = 84.0$, $SE = 3.1$ for Blocks 1 and 2, respectively). Chance performance is 1 out of 5 response categories, or 20% correct. Because the symptom–disease relationships were deterministic, the per-

Table 4
Percentages of Correct Diagnoses on the Last Two Occurrences of Each Trial Type in the Learning Stage of Experiment 2

Trial type	% correct
Block 1	
AB	91.4
B	97.7
DE	89.1
E	95.3
M	93.4
Block 2	
AC	82.8
C	85.2
DF	73.4
F	94.5
M	84.0

Note. Chance performance is 20%. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms.

centage of correct responses is of course greater than in Experiment 1.

The main results are participants' contingency judgments in the two trial order conditions. Mean ratings from the inh–con condition ($M = 69.4$, $SE = 3.6$) were higher than ratings from the con–inh condition ($M = 19.7$, $SE = 7.4$). A t test showed that the difference was statistically significant, $t(31) = 7.68$.

The results revealed that contingency judgments were again affected by trial order, at variance with the predictions of the PCM-ALG. Moreover, the specific trial order effect found is predicted by the Rescorla–Wagner theory.

In what follows, we undertake an evaluation of whether participants' estimations of the target relationships evolved across trials in the manner predicted by the theory. The differences between the proportion of times the target diseases were diagnosed on the relevant trial types (*AB* vs. *B* and *AC* vs. *C*, and *DE* vs. *E* and *DF* vs. *F*) serve this purpose. These differences are shown in the continuous lines of Figure 2. Again, a general parallel between participants' performance and the associative theory's predictions can be seen. A 2 (order: con–inh vs. inh–con) × 10 (blocks) within-subjects ANOVA was performed on these differences. The blocks variable included two trial presentations of each target trial type. Both within-subjects variables, order and blocks, were significant, $F(1, 31) = 26.28$, $MSE = 0.17$, and $F(9, 279) = 2.25$, $MSE = 0.08$, $p < .05$, respectively. The interaction between the variables also was significant, $F(9, 279) = 227.24$, $MSE = 0.12$. Four nonorthogonal planned contrasts served to evaluate the model's main predictions. In the con–inh condition, the difference after the fifth block of trials was significantly greater than after the first block, $F(1, 31) = 119.77$, $MSE = 0.12$. Similarly, the difference after the last block of trials was smaller than after the fifth block of trials, $F(1, 31) = 450.59$, $MSE = 0.12$. In the inh–con condition, the difference after the fifth block of trials was smaller than after the first block, $F(1, 31) = 166.37$, $MSE = 0.12$. In addition, the difference after the last block of trials was greater than after the fifth block, $F(1, 31) = 507.11$, $MSE = 0.12$. Thus, the main characteristics of the model's predictions were well illustrated by the participants' performance across trials.

One discrepancy between the model's predictions and participants' judgments is that the former predicts a negative terminal associative strength for cues in the con–inh condition, whereas judgments had a mean positive value of 19.7. How can we explain this discrepancy? Participants may have had difficulties with the use of the response scale, which means that they could not clearly distinguish between its negative half and values close to zero. In fact, participants only used the negative half of the scale on 12 out of 64 possible occasions (2 per participant) in which the relationship was negative during Block 2. In line with this, participants were informally interviewed at the end of the task, and some of them indicated difficulties in distinguishing between negative judgments and ones close to zero. This discrepancy may then be understood in terms of a performance deficit rather than an associative one. The general agreement shown in Figure 2 between the model's predictions and participants' performance also speaks in favor of this possibility, and the monotonic relationship assumed by the Rescorla–Wagner theory between the cues' associative strengths and participants' judgments is reinforced in the results.

The absolute difference between judgments in the two trial order conditions in Experiment 1 was 7.0 (80.6 vs. 73.6), whereas in Experiment 2 it was 49.7 (69.4 vs. 19.7). Plainly, in absolute terms, the trial order effect found in Experiment 2 was of a much greater empirical magnitude than the one found in our previous experiment, as predicted by the associative theory. Given the relevance of trial order effects from a theoretical perspective, having found an empirically robust trial order effect is of considerable interest.

We have argued that the trial order effects we have obtained are at variance with the PCM-ALG, but of course it is possible to think of various ways in which the data and theory might be reconciled. Therefore, in Experiments 3 and 4 we explored two straightforward possibilities for explaining the results without contradicting the PCM-ALG's predictions. We concluded, however, that neither of these possibilities is viable.

Experiment 3

A simple explanation of the effects of trial order found in Experiments 1 and 2 is that the Block 2 trial types led to forgetting or unlearning of the Block 1 types. Clearly, as far as the target cues are concerned, the patterns of association presented in the two blocks are largely contradictory, and so it is plausible to imagine that some degree of forgetting of the earlier trial associations is caused. Indeed, evidence from verbal list-learning experiments using an *A–B*, *A–C* design suggests that the number of trials presented in Block 2 may have been sufficient to lead to very substantial levels of forgetting (e.g., Barnes & Underwood, 1959). If, at the time of making their contingency ratings, participants had no memory of the Block 1 trial types, then the trial order effects

observed would be expected even on the PCM-ALG, because the contingency during Block 2 between cue *A* and outcome O_1 was lower than the contingency between cue *D* and outcome O_2, both in Experiments 1 and 2. This hypothesis is especially plausible if we take into account the likelihood that participants presumably cannot keep track of all of the information presented during the task because of an obvious limited capacity of information processing. Provided that participants have unlearned all of the Block 1 trials types, the PCM-ALG can give an adequate account of the results (see Appendix). Although the following analysis only refers to Experiment 2, a similar analysis can be performed for Experiment 1.

The PCM-ALG's predictions are modified if Block 1 trials are not included in the computations of the conditional contrasts. It can easily be seen from Table 3 that there is a negative contingency between cue *A* and outcome O_1 and a positive one between cue *D* and outcome O_2 if only Block 2 trials are considered. Hence, different ratings are predicted for the two trial order conditions. Note, however, that if just one trial of each type from Block 1 is recalled and included in the calculation, the model predicts no trial order effect, as it is probabilities rather than event frequencies that matter for the PCM-ALG. In this case, the set of probabilities would be identical for *A* and *D* and, hence, the same conditional contrasts would be computed yielding the same contingency values. Thus, the model would not be able to account for the trial order effect previously found.

The aim of Experiment 3 was, thus, to evaluate the viability of this forgetting hypothesis. For this, a test stage was included after participants had made their contingency judgments. In this test stage, two trials of each trial type (from Blocks 1 and 2; see Table 3) were re-presented without any feedback. The key issue is whether participants can remember the correct responses for the Block 1 trial types. If they can, then the forgetting hypothesis becomes unviable. However, if participants have forgotten or unlearned the Block 1 trial types, they will obviously make more errors on test trials from Block 1 than from Block 2. In addition, another prediction can be derived from this hypothesis. For example, if participants are presented with symptom *C* in the test stage, we will expect them to diagnose disease O_1 with greater probability than they do when presented with symptoms *A* and *C* combined. If participants have unlearned the Block 1 trial types, then they should apply the negative contingency experienced for symptom *A* in Block 2 to the Block 1 test trials as well because this is the only source of evidence available to them concerning symptom *A*'s predictive role. Thus, participants will reverse their diagnoses made during the first half of the learning stage itself. Although participants should be more inclined to diagnose O_1 on an *AB* trial than on a *B* trial during Block 1 trials, they should (on the forgetting hypothesis) more often diagnose O_1 on *B* rather than on *AB* trials during the test stage (i.e., after experiencing a negative contingency for symptom *A* and disease O_1 in Block 2). A similar argument can be made for symptom *D*, with the opposite change of predictions. In this case, participants should be more inclined to diagnose O_2 on *DE* than on *E* trials during the test stage, even though the opposite diagnoses should have been made in Block 1 itself.

To sum up, according to the forgetting hypothesis, individuals should make more errors on the test trials from Block 1 than from Block 2. Also, people should make more O_1 diagnoses in the test stage on *B* trials than on *AB* trials and, conversely, they should make more O_2 diagnoses on *DE* than on *E* trials, despite the fact that the opposite pattern of diagnoses should have been made in Block 1 of the learning stage.

Method

Participants and apparatus. A total of 20 psychology students from Universidad de Málaga volunteered for this experiment. They had not participated in the previous experiments. The same computer and similar questionnaires were used as in Experiments 1 and 2.

Procedure. Procedural details were as in the previous experiments. The learning stage was identical to the one described for Experiment 2 (see Table 3). Thus, two sets of con–inh and inh–con trials were again programmed, and contingency judgments were also required at the end of this learning stage.

After participants had made their ratings, a test stage was presented consisting of 32 trials. Each of the 16 different trial types from the learning stage was presented twice in a random order. Unlike in the learning stage, participants were not given any feedback in the test stage about their diagnoses.

Results and Discussion

Table 5 shows the percentage of correct responses across participants for the last two trials of each trial type at the end of each block of the learning stage. These mean percentages show that by the end of each block participants had a good knowledge of the programmed relationships ($M = 92.5$, $SE = 2.9$ and $M = 81.3$, $SE = 4.8$ for Blocks 1 and 2, respectively), chance performance being 20% correct.

Table 5
Percentages of Correct Diagnoses on the Last Two Occurrences of Each Trial Type in the Learning Stage of Experiment 3

Trial type	% correct
Block 1	
AB	91.3
B	97.5
DE	93.8
E	87.5
M	92.5
Block 2	
AC	78.8
C	78.8
DF	70.1
F	97.5
M	81.3

Note. Chance performance is 20%. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms.

Participants' contingency judgments after the learning stage replicated our previous findings. Ratings from the inh–con condition were higher than ratings from the con–inh condition ($M = 45.0$, $SE = 6.7$ and $M = 18.2$, $SE = 9.0$, respectively; see Table 3). A t test showed that the difference was statistically significant, $t(19) = 3.8$. At variance with the PCM-ALG's predictions but consistent with those of the Rescorla–Wagner theory, judgments were sensitive to trial order.

The main interest of this experiment was to evaluate participants' performance during the test stage. Table 6 shows the mean percentages of correct responses for each trial type during this test. As can be seen, the mean percentages of correct responses for those trial types originally presented during Blocks 1 and 2 of the learning stage were $M = 71.6$, $SE = 4.1$ and $M = 70.3$, $SE = 3.2$, respectively. These two mean percentages did not differ statistically, $t(19) = 0.32$. Thus, participants seemed to remember with a similar degree of accuracy trials from both blocks of the learning stage, which is at variance with the forgetting hypothesis. According to this hypothesis, participants should have made more errors on Block 1 trials, as these trials should have been displaced from memory by the more recent information, that is, the Block 2 trials.

Moreover, according to the forgetting hypothesis participants in the test stage should have reversed the diagnoses they made during the first half of the learning stage, as described above. Specifically, the proportion of O_1 diagnoses on *AB* trials should be greater than on *B* trials during Block 1 of the learning stage but lower during the test stage. Correspondingly, the proportion of O_2 diagnoses on *DE* trials should be lower than on *E* trials during Block 1 of the learning stage but higher during the test stage. Table 7 displays the proportion of these diagnoses for each trial type during Blocks 1 and 2 of the learning stage and in the test stage. The table shows that, contrary to the PCM-ALG, responding in the test stage was consistent with responding at the end of each learning stage. A 2 (order: con–inh vs. inh–con) × 2 (stage: block 1 vs. test) × 2 (type of cue: compound vs. single) ANOVA was performed on these proportions. The main effects of the within-subjects variables order and stage were significant, $F(1, 19) = 8.68$, $MSE = 0.08$, and $F(1, 19) = 12.49$, $MSE = 0.02$, respectively. The interactions Order × Stage, Order × Type of Cue, and Order × Stage × Type of Cue were also significant, $F(1, 19) = 15.95$, $MSE = 0.03$; $F(1, 19) = 224.84$, $MSE = 0.08$; and $F(1, 19) = 35.78$, $MSE = 0.03$, respectively. The analyses to evaluate the forgetting hypothesis were, though, simple effect contrasts. At variance with this hypothesis, the contrasts revealed that the proportion of O_1 diagnoses was higher on *AB* trials than on *B* trials, both in Block 1 and in the test stage, $F(1, 19) = 671.58$, $MSE = 0.01$, and $F(1, 19) = 35.58$, $MSE = 0.06$, respectively. Similarly, the proportion of O_2 diagnoses on *DE* trials was lower than on *E* trials, both in Block 1 and in the test stage, $F(1, 19) = 198.75$, $MSE = 0.03$, and $F(1, 19) = 35.77$, $MSE = 0.09$, respectively. Despite not having forgotten the information provided during Block 1, an inspection of the data reveals that participants were at the same time sensitive to the negative relationship arranged during Block 2 of the learning stage between *A* and O_1 and the positive one between *D* and O_2. Participants only chose disease O_1 on 23.8% of *AC* trials, whereas they chose it on 70.0% of *C* trials. In contrast, participants chose disease O_2 on 48.8% of *DF* trials, whereas they only chose it on 6.3% of *F* trials, consistent with the contingencies programmed in Block 2.

Table 6
Percentages of Correct Diagnoses of Each Trial Type in the Test Stage of Experiment 3

Trial type	% correct
Block 1	
AB	48.8
B	93.8
DE	65.0
E	78.8
M	71.6
Block 2	
AC	72.5
C	70.0
DF	48.8
F	90.0
M	70.3

Note. Chance performance is 20%. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms.

Table 7
Percentages of Diagnoses of the Target Diseases Across the Different Trial Types During the Learning (Last Two Occurrences) and Test Stage of Experiment 3

Trial type	% correct
Learning stage	
Block 1	
AB	91.3
B	1.3
DE	3.8
E	87.5
Block 2	
AC	11.3
C	78.8
DF	70.0
F	1.3
Test stage	
Block 1	
AB	48.8
B	2.5
DE	21.3
E	78.8
Block 2	
AC	23.8
C	70.0
DF	48.8
F	6.3

Note. For trial types *AB*, *B*, *AC*, and *C*, the predictions refer to disease O_1. For *DE*, *E*, *DF*, and *F*, the predictions refer to disease O_2. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms.

In sum, these results are at variance with the forgetting hypothesis. Participants' performance during the test stage was not exclusively sensitive to the contingencies presented during Block 2, as it would be if Block 1 trials had been forgotten or unlearned. Thus, the trial order effect found is inconsistent with the PCM-ALG's predictions.

What are the implications of these test data for the associative theory? Although the trial order results obtained thus far in the present series of experiments are predicted by the Rescorla–Wagner theory, the test stage results from the present experiment are deeply problematic for the associative account. As can be seen in Figure 2, the model predicts a strongly negative and positive associative strength for cues *A* and *D*, respectively, at the end of Block 2. Thus, participants' performance in the test session should have been sensitive to these final predictive values of both target cues. The negative associative strength of cue *A* (equal at asymptote to $-\lambda$) should have led participants to predict the absence of disease O_1 on *AB* trials, at least, just as on *B* trials. Also, the positive associative strength of cue *D* (equal at asymptote to λ) should have led participants to predict the presence of disease O_2 on *DE* trials, at least, just as on *E* trials. But neither of these predictions was fulfilled, as shown above. From the point of view of the model, the associative strengths of *A* and *D* in the test stage are paradoxical: *A* increases the proportion of outcome predictions when compounded with *B* but decreases it when compounded with *C*, and similarly for *D* when compounded with *F* and *E*.

In fact, in this experiment the Rescorla–Wagner theory predicts the well-known phenomenon of *catastrophic forgetting* (Kruschke, 1993; Lewandowsky, 1991; McCloskey & Cohen, 1989; Ratcliff, 1990; Sloman & Rumelhart, 1992). As a result of learning the new relationships programmed between cues and outcomes during Block 2, participants should have unlearned all previous relationships. Contrary to this prediction, participants' diagnostic responses during the test session showed a similar degree of accuracy for trial types originally presented in Blocks 1 and 2 of the learning stage.

The problem of catastrophic forgetting is not restricted to the Rescorla–Wagner theory but also represents a limitation for other associative or connectionist models. For instance, backpropagation networks using the generalized delta rule as a learning algorithm (Rumelhart, Hinton, & Williams, 1986) predict the same catastrophic forgetting effect. In fact, this effect was originally described in relation to such connectionist models (McCloskey & Cohen, 1989). By way of illustration, we conducted a series of simulations using backpropagation networks as described by McClelland and Rumelhart (1988). The training patterns coincided with the training trials participants received during the learning stage of Experiment 3 and were also distributed in two blocks within each trial order condition (con–inh: $AB \rightarrow O_1$, $B \rightarrow$ no *O* followed by $AC \rightarrow$ no *O*, $C \rightarrow O_1$; inh–con: $DE \rightarrow$ no *O*, $E \rightarrow O_2$ followed by $DF \rightarrow O_2$, $F \rightarrow$ no *O*). Different parameters and network architectures (different numbers of hidden units) were tested in these simulations but all of them showed the basic catastrophic forgetting effect; thus, only the details of one of them are described. The outcome of this simulation is presented in Figure 3. A three-layered network was used consisting of six input units corresponding to the

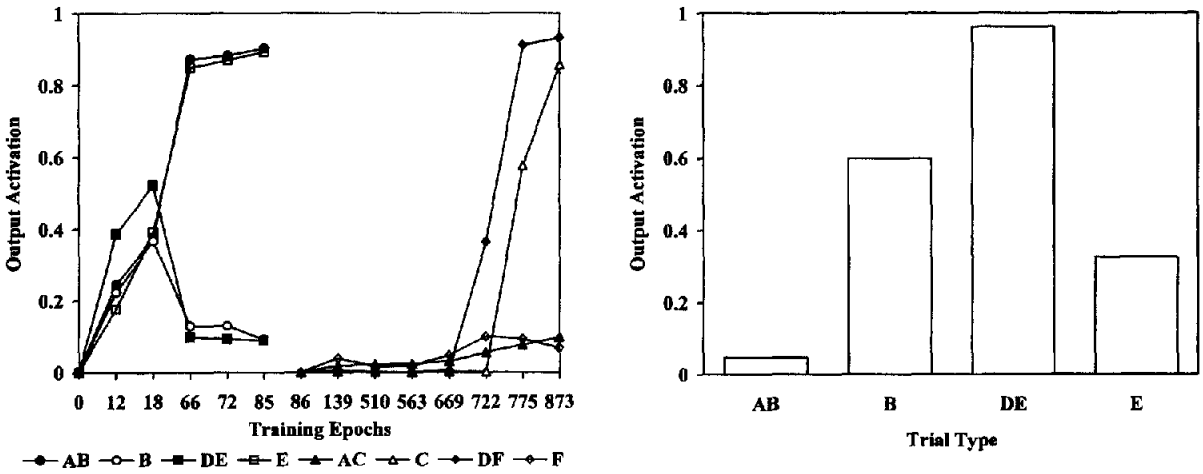

Figure 3. Left: changes in output unit activations of a back-propagation network across training epochs for the different trial types of Experiment 3. An epoch consists of one presentation of each trial type. The activation of output unit O_1 is shown on *AB*, *B*, *AC*, and *C* trial types. The activation of output unit O_2 is shown on *DE*, *E*, *DF*, and *F* trial types. The learning rate was set to .2 and the momentum parameter to .9. The remaining free parameters were set to default values. Input biases were adjusted to all hidden and output units. The training of the network continued until a learning criterion had been met (*tss* = .04). Right: output unit activations after the training phase had been completed for the different trial types from Block 1. The network's performance during the test stage showed that it had catastrophically forgotten the relationships learned during the first block of the learning stage.

six cues (symptoms), 10 hidden units to allow an internal representation of the input information to be formed, and two output units corresponding to the two outcomes (diseases) used in the training stage. "No disease" was coded as an output of zero on both output units. As can be seen, in Block 1 of the learning stage the activation values of the O_1 and O_2 output units increased across trials up to asymptote on *AB* and *E* trials, respectively. In contrast, on *B* and *DE* trials, the activation values of both output units remained low across this first block of trials. During Block 2, the activation values on the O_1 and O_2 output units on *C* and *DF* trials, respectively, increased across trials towards asymptotic values, but only if a very large number of trials was presented. The activation values of the output units were low on *AC* and *F* trials within the second block of trials.

The main results are the activation values of the output units on the different trial types in the test stage (see right panel of Figure 3). The activation of the output unit corresponding to outcome O_1 was .05 on *AB* trials, whereas on *B* trials it was .60 on a scale of 0 to 1. Therefore, the network predicts outcome O_1 more often on *B* than on *AB* trials, despite the $AB \rightarrow O_1$, $B \rightarrow$ no *O* training presented in Phase 1. Similarly, the activation of the output unit corresponding to outcome O_2 was .96 on *DE* trials, whereas it was .33 on *E* trials. Therefore, the network predicts outcome O_2 more often on *DE* than on *E* trials. This pattern of results showed that the network performed during the test stage according to the contingencies it had learned during Block 2 of the learning stage and, thus, the relationships programmed during Block 1 were catastrophically forgotten.

Despite these falsified predictions, there are several possible solutions to the catastrophic forgetting problem within the framework of associative models (Lewandowsky, 1994). In the General Discussion section we return to this issue and explore one particular way of effectively addressing this problem. The solution is based on the creation of a semi-distributed representation of the input information by the assignment of configural coding units to each input pattern together with the inclusion of a generalization mechanism based on input similarity (Pearce, 1994).

Experiment 4

Returning to the PCM-ALG, the results from our previous experiments present difficulties for this model. Experiment 3 considered a simple way in which the model's predictions and the data could be brought into line, but this proved unviable. In Experiment 4, we explored a somewhat more subtle approach that may make the trial order effect found still compatible with the PCM-ALG.

In Experiments 1–3, trials were always presented as two very distinct blocks. Although these blocks were presented as a continuous sequence of trials, participants surely realized that contingencies changed dramatically in the middle of the learning stage. Accordingly, they may have reasoned that only the more recent trials should be considered when making their contingency judgments. Then, participants may well have been operating with a moving window that only extends backwards a certain number of trials. To put it another way, only the more recent trial types may have been included in the window, and only the contents of the window contributed to the computation of contrasts. The remaining trial types are simply excluded from the focal sets over which contrasts are computed. Such a mechanism allows the results of the previous experiments to be accommodated by the PCM-ALG, because across Block 2 the conditional contingency for cue *D* is greater than that for cue *A*, as described in Experiment 3. Moreover, participants do not need to unlearn the Block 1 relationships on this account, as they potentially can access all of the trial types that have been observed. It is simply that certain trial types have been excluded, though not forgotten, from the focal sets. Hence, correct responding on the Experiment 3 test trials would still be possible.

To control the potential effects of this moving window idea on the PCM-ALG's predictions, trials were not presented in two separate blocks in the present experiment, but were instead presented in a fully intermixed way. Specifically, in the con–inh condition, participants saw that the positive contingency trials ($AB \rightarrow O_1$, $B \rightarrow$ no *O*) predominated at the start of the learning stage, whereas the negative contingency trials ($AC \rightarrow$ no *O*, $C \rightarrow O_1$) were more numerous towards the end. Nevertheless, participants could also see some negative and positive contingency trials at the start and end of the learning stage, respectively. On the other hand, in the inh–con condition, negative contingency trials ($DE \rightarrow$ no *O*, $E \rightarrow O_2$) were predominant at the start of the learning stage, whereas positive contingency ones ($DF \rightarrow O_2$, $F \rightarrow$ no *O*) occurred more often at the end of it. Again, participants could see both positive and negative contingency trials at the start and end of the learning stage. Table 8 displays the actual design used during this learning stage.

Although the PCM-ALG anticipates the trial order effects observed in the previous studies if this moving-window hypothesis is considered, it predicts that participants' judgments in the present experiment should not be affected by trial order. The modifications introduced in the distribution of trials across the task do not alter the evidence concerning the $A \rightarrow O_1$ and $D \rightarrow O_2$ relationships, which is identical for both trial order conditions. As every trial type can occur at

Table 8
Design of Experiment 4

Trial order condition	No. of trials within blocks 1	2	3	4	Trial types	Test cue	Judgment
Con–inh	4	3	2	1	$AB \rightarrow O_1$	$A \rightarrow O_1$?	$M = 39.7$
	4	3	2	1	$B \rightarrow$ no *O*		
	1	2	3	4	$AC \rightarrow$ no *O*		
	1	2	3	4	$C \rightarrow O_1$		
Inh–con	4	3	2	1	$DE \rightarrow$ no *O*	$D \rightarrow O_2$?	$M = 52.9$
	4	3	2	1	$E \rightarrow O_2$		
	1	2	3	4	$DF \rightarrow O_2$		
	1	2	3	4	$F \rightarrow$ no *O*		

Note. Con = contingent; inh = inhibitory. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms; O_1 and O_2 are diseases 1 and 2, respectively; no *O* = no disease.

any point throughout the learning stage, the conditional contrasts yield the same values for both trial order conditions, regardless of how narrow or wide this temporal window is (see the Appendix). Imagine that the window is sufficiently restricted to include only trials from the last quarter of the learning stage. Because all trial types occur within this last quarter (although in different proportions), the conditional contingencies are identical for target cues A and D and identical to the previous analysis. The focal sets on which the contrasts are computed are independent of trial frequency. For example, let us examine the conditional contrasts for cue A (con–inh condition). With regard to the computation of $\Delta P_{\overline{C}}$ $[P(O_1/A \cdot C) - P(O_1/\overline{A} \cdot C)]$, trial types including cue C were initially infrequent but became predominant by the end of the learning stage. Nevertheless, this contrast yields a value of -1.0. With regard to the computation of ΔP_C $[P(O_1/A \cdot \overline{C}) - P(O_1/\overline{A} \cdot \overline{C})]$, this can only be computed if $AB \rightarrow O_1$ trial types are included within the window. Note, however that the inclusion of just one trial of this trial type will be sufficient to allow this contrast to be computed, and it yields the same value as if many of them are included, that is, a value of 1.0. For situations in which all conditional probabilities are either 0 or 1.0, it makes no difference whether a single trial or many trials are observed of a given type: The conditional probability is the same after one trial as after many. Obviously, the unconditional ΔP measure yields different values for the two trial order conditions if only the last quarter of the training trials are included. However, this measure has already been rejected as a theory of human contingency detection on empirical and normative grounds for situations involving a varying context (see above; Cheng & Holyoak, 1995; Shanks, 1991; Wasserman et al., 1993).

In contrast, the Rescorla–Wagner theory predicts the same recency–density effect on participants' ratings as described above. However, as a side effect, this new trial type arrangement entails a major modification in the nature of the task that severely affects some of the model's predictions. The trial types create a nonlinearly separable discrimination problem of the sort that is well known to be problematic for the Rescorla–Wagner theory (see Shanks, 1995b). This is because there is no set of associative weights that allow participants to make correct predictions for all trial types. For example, within the con–inh condition, participants saw $AB \rightarrow O_1$, $B \rightarrow$ no O, $AC \rightarrow$ no O, and $C \rightarrow O_1$ trials. If participants are to predict the absence of the outcome on B trials, cue B should have a zero associative strength and, thus, cue A an associative strength of λ if O_1 is to be predicted on AB trials. Correspondingly, if participants predict O_1 on C trials, cue C should have an associative strength of λ and, thus, cue A an associative strength of $-\lambda$ if no outcome is to be predicted on AC trials. However, cue A cannot simultaneously have a positive and a negative associative strength for the same outcome. Thus, participants should not be able to predict the correct outcomes for the different trial types.

Finally, a test stage was also included after participants had made their contingency judgments. The purpose of this was to further evaluate whether participants' performance suffers from catastrophic forgetting. Because trials have been interleaved to some extent across the learning stage, a catastrophic forgetting effect is now less likely, as participants had the chance to see all trial types at every point during the task.

Method

Participants and apparatus. A total of 44 psychology undergraduates from Universidad de Málaga took part in this experiment on a voluntary basis. They had not participated in the previous experiments. The same apparatus and similar questionnaires were used as in the previous experiments.

Procedure. The task and the rest of the procedural details were as described previously. Again, two trial order conditions were presented, con–inh and inh–con. The learning stage consisted of 160 trials and it was divided into four blocks. Within Block 1, four trials of each of the following trial types were presented: $AB \rightarrow O_1$, $B \rightarrow$ no O and $DE \rightarrow$ no O, $E \rightarrow O_2$. In addition, one of each of the following trial types was also presented: $AC \rightarrow$ no O, $C \rightarrow O_1$ and $DF \rightarrow O_2$, $F \rightarrow$ no O. During Block 2, three and two trials of the former and latter trial types were presented, respectively, whereas in Block 3, two and three trials of the former and latter trial types were presented. Finally, in Block 4, one and four trials of the former and latter trial types were included, respectively. The counterbalancing procedure was similar to that for previous experiments. Participants' judgments were also required at the end of the learning stage, and then the test stage took place.

Results and Discussion

The percentage of correct diagnoses for the last two trials of each trial type from the learning stage is displayed in Figure 4. These mean percentages, though lower than in previous experiments, still showed that participants correctly learned the programmed contingencies. As can be seen, however, the percentages varied greatly across different trial types (from 54.0 to 90.9). Thus, a 2 (trial type: predominant in the first half vs. predominant in the second

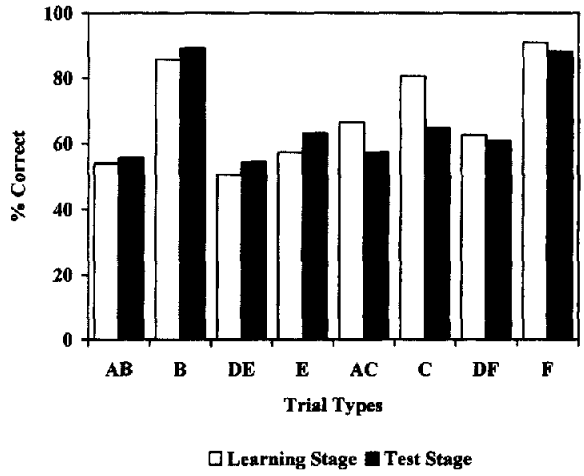

Figure 4. Percentages of correct diagnoses of each trial type in the learning (last two occurrences) and test stage of Experiment 4. Chance performance is 20%.

half of the learning stage) × 2 (type of cue: single vs. compound cue) within-subjects ANOVA was conducted on these percentages to establish the source of this variability. The main effects of both variables, trial type and type of cue, were significant, $F(1, 43) = 28.71$, $MSE = 0.03$, and $F(1, 43) = 56.41$, $MSE = 0.03$, respectively. The mean percentage of correct responses for those trials predominant in the first half was lower than the percentage for trials predominant in the second half of the learning stage ($M = 61.9$, $SE = 2.4$ and $M = 75.1$, $SE = 2.9$, respectively). The fact that trials were presented in a 1:4 ratio in the last block of the learning stage may explain why these percentages differed. Also, the mean percentage of correct diagnoses for single cues ($M = 78.7$, $SE = 1.9$) was higher than the percentage for compound cues ($M = 58.4$, $SE = 3.3$). Given that compound cues included the target cues, whose predictive value had changed across the task, it is not surprising that participants responded more accurately to single cues.

Mean contingency ratings from the inh–con condition were higher than mean ratings from the con–inh condition ($M = 52.9$, $SE = 5.3$ and $M = 39.7$, $SE = 7.1$, respectively). A t test revealed that the difference was significant, $t(43) = 3.27$. Again, participants' judgments were sensitive to trial order in a situation in which the PCM-ALG predicts no such effect. In this case, trials were not distributed in two distinct blocks, yet ratings were affected by trial order, which speaks to the sensitivity of the contingency detection mechanism to this effect and also to its empirical robustness.

Figure 4 also shows the mean percentage of correct responses for each trial type in the test stage. These results showed that participants' performance was practically as accurate within the test stage as it was at the end of the learning stage ($M = 66.7$ and $M = 68.5$, respectively). Though the contingencies had smoothly changed in the learning stage, participants made similar numbers of errors on those test trial types more numerous in the first half than on test trial types predominant in the second half of the learning stage ($M = 65.6$, $SE = 2.6$ and $M = 67.8$, $SE = 2.6$, respectively), and the small difference was not statistically significant, $t(43) = 0.78$. This comparison replicates the basic finding observed in the test stage of Experiment 3.

The pattern of results obtained is problematic for the PCM-ALG, even when a hypothetical temporally restricted window is adopted. Different ratings were obtained in the two trial order conditions. The main problem for the theory is that trial types predominant during the first half of the training session cannot be excluded from the temporal window on which contrasts are supposed to be computed, as these trial types are as recent as any other. Hence, no trial order effect can be expected on the PCM-ALG.

On the other hand, though the trial order effect found is predicted by the Rescorla–Wagner theory, some of the results described thus far are at variance with this associative account. First, participants did not make more errors in the test stage on trials predominant early compared to those predominant late in training. The theory predicts this difference for the same reason that it also predicts the comparable catastrophic forgetting effect in Experiment 3, though to a lesser extent due to the intermixture of trial types. Second, and more directly, participants' performance was well above chance at the end of the learning stage. According to the theory, the learning task should not even be solvable, as described before. In the next section, some possible solutions to these various problems are discussed.

General Discussion

The series of experiments reported here has provided clear evidence about the effect of the order in which trials are presented on judgments of predictive relationships. Specifically, our studies provide the first demonstration of recency–density effects in human contingency judgment. Experiment 1 revealed that participants' contingency judgments were lower in a con–non condition than in a non–con one, implying that later information exerted a greater impact on judgments than earlier information. Experiments 2–4 replicated this basic result and extended it to situations in which more radical changes in the programmed contingencies were made. When the contingencies changed from being positive to negative or vice versa, the results provided evidence of a more substantial trial order effect. In addition, Experiments 3 and 4 included a test stage to evaluate whether participants forgot trial types presented early during the learning stage. The results showed that this was not the case, as a similar number of correct diagnostic responses was made during this test stage for all trial types. In addition, Experiment 3 found that participants' diagnostic choices during this test stage for trial types originally presented within the first half of the learning stage were not affected by the later exposure to very different contingencies. Moreover, Experiment 4 showed that the trial order effects were not due to the fact that in Experiments 1–3 we presented the learning stage trials in two separate blocks. A trial order effect was observed even in a situation in which trials were interleaved in a single block, and this finding contradicts the idea that individuals compute contrasts across the contents of a restricted window of trials. Hence, our results consistently revealed that later and more dense trials had the greatest impact on individuals' judgments. Recency effects are observed in a number of other judgment tasks (Hogarth & Einhorn, 1992).

With regard to the scope of the data we have reported, we think that these trial order effects should be taken as an important empirical phenomenon for which models of the psychological processes involved in contingency detection must offer an account.

Implications for the PCM

As described in the introduction, we have been concerned with testing an algorithmic-level version of the PCM. This model, PCM-ALG, makes clear-cut predictions about the role of trial order on contingency judgments. Given that trial order does not alter the various probabilistic contrasts that people compute, it should not modify their contingency ratings. In this sense our results clearly contradict the model's predictions.

In our applications of the model, we have assumed that participants' behavior is appropriately modeled by the computation of the conditional contingency between the target symptom and disease on the basis of the probability of the disease in the presence and absence of the symptom. However, as mentioned in the introduction, Waldmann and Holyoak (1992) have suggested that in tasks such as medical diagnosis participants may impose a causal model on the situation such that the symptoms are interpreted as effects of the disease, and may then base their contingency judgments on a calculation of the conditional contingency between the target disease and symptom on the basis of the probability of the symptom in the presence and absence of the disease. On this basis, the calculations of conditional contrasts are somewhat different from the ones we have presented. But it is important to note that this causal model theory suffers exactly the same problems as the version we have considered. Because the same trial types are presented in the two trial order conditions, the disease–symptom contingencies are identical for diseases 1 and 2 and no difference in judgments is predicted. Moreover, recent evidence has questioned other aspects of the causal model theory (Matute, Arcediano, & Miller, 1996; Price & Yates, 1995; Shanks & López, 1996; but see Waldmann & Holyoak, 1997).

The failure of the PCM-ALG to explain the results comes about, we think, as a consequence of the sort of questions the model is meant to address. A normative analysis prescribes the environmental conditions that must be met in order for the individual to assert that a contingency exists in a given situation. This normative discourse, namely contingency theory, provides an independent measure of the circumstances in which two events can be regarded as being truly related. The PCM-ALG incorporates such a normative analysis, on the basis of philosophical and statistical considerations, and assumes that people's performance is reasonably well adjusted to this normative standard. As a psychological model, the PCM-ALG specifies a mental algorithm to exclusively compute such norms. Thus, the application of such an algorithm will ensure that people will only detect a relationship between events when this relationship truly exists.

Normative considerations constitute, no doubt, an appropriate source of constraints for any psychological theory meant to account for behavior. However, there are further constraints to which theories must attend and that are beyond the scope of a normative framework (see López, Cobos, Caño, & Shanks, in press). These constraints are of a psychological rather than a normative nature. Specifically, our experiments have shown that people are sensitive to how the relevant information is distributed across time. It may well be the case that such distribution is not relevant for a statistical or normative definition of what a relationship is, but it is surely the case that it is psychologically relevant, as our results suggest. Similarly, other investigations have also shown that people's performance is sensitive to how trials are distributed across the task with regard to cue competition effects (Davey & Singh, 1988; Lovibond, Siddle, & Bond, 1988; Martin & Levey, 1991).

Nevertheless, this further constraint does not in our view necessarily constitute an example of irrational behavior or behavior poorly adapted to a changing environment. For example, consider the experimental conditions in Experiment 2. In the con–inh condition, participants saw $AB \rightarrow O_1$, $B \rightarrow$ no O, followed by $AC \rightarrow$ no O and $C \rightarrow O_1$ trials, whereas in the inh–con condition they saw $DE \rightarrow$ no O, $E \rightarrow O_2$, followed by $DF \rightarrow O_2$ and $F \rightarrow$ no O trials. Though statistically or normatively the $A \rightarrow O_1$ and $D \rightarrow O_2$ relationships are identical, the situation is not obviously the same in the two conditions for an individual whose objective is to maintain an updated knowledge base about the predictive value of a cue on a moment-by-moment basis. The distribution of trials across the task does not affect the statistical relationship itself, but a system sensitive to changes in the distribution of trials has an obvious adaptive advantage. In this sense, the PCM-ALG as currently formulated lacks a mechanism able to manifest such sensitivity.

The diagnosticity of trial order effects to distinguish between the contingency-based and the associative model crucially depends on the assumption that such effects are due to the operation of mental processes specifically related to contingency detection. One alternative explanation still compatible with the PCM-ALG's predictions was explored in Experiments 3 and 4. The results showed that the trial order effect was not due to the forgetting or unlearning of those trial types presented during the first half of the learning stage. There is, however, another alternative interpretation of the trial order effects found in Experiments 1–3 also compatible with the PCM-ALG's predictions, namely, that they arise because of a temporal restriction of the trials on which the conditional contrasts are computed. According to this hypothesis, conditional contrasts would be computed on those trials presented most recently during the learning stage, with the initial trials being effectively excluded. The results obtained in Experiment 4, though, were at variance with the PCM-ALG even if participants were operating on a temporally restricted window. Regardless of the size of this moving window, the distribution of trial types ensured that every trial type should have been included in it and, hence, no trial order effect should have been observed, as in this case the conditional contrasts yield identical values as in Experiment 2. In terms of the PCM-ALG, the inclusion of a single trial of those trial types predominant at the start of the learning stage within the final window will be sufficient to allow the same conditional contrasts to be computed as if all trial types were evenly included within such a window.

In line with our results, it seems that it is the relative density of trial types at the end of a learning phase that is the crucial factor that leads to trial order effects on contingency judgments. Accordingly, it is very difficult to see how a normative model that assigns no role to the actual distribution across time of the event frequencies that contribute to conditional probability estimates can explain this basic effect. In a way, it is the tribute that such a model has to pay to the normative analysis that inspires it (i.e., contingency theory) that determines how an objective relationship between events is defined in statistical or probabilistic terms. It seems unlikely that the basic units of analysis in the PCM and the PCM-ALG, namely accurate estimation of conditional probabilities, are appropriate at the algorithmic level of psychological analysis.

Of course, the PCM-ALG is only one of many possible

implementations of the computational-level theory (the PCM), and we must await future theoretical development of this theory. Nevertheless, it is hard to see how any model that computes conditional contrasts, as the PCM requires, can account for trial density effects. We view Cheng's (1997) recent Power PC model in a similar light. This computational-level model, a development of the PCM, has not yet been implemented as a processing model, but as with the PCM, it is hard to see how it can account for trial order effects on the basis of conditional probabilistic contrasts.

Implications for the Rescorla–Wagner Theory

The sensitivity of people's performance to the distribution of trial types across time can be understood in terms of an associative account such as the Rescorla–Wagner theory. Basically, two conditions are met by the associative theory that allow it to predict such a trial order effect. First, the learning algorithm is sensitive to event frequencies within a fixed period of time. Second, the algorithm is flexible enough so as to adapt on a moment-by-moment basis to the input information and, accordingly, be capable of generating an updated response at any point across the task.

Another aspect of our results that also agrees with this particular associative account is the pattern of evolution across trials of participants' diagnostic responses. Specifically, the differences between the proportion of times the target diseases were predicted on compound and single cue trial types allowed us to indirectly track how the predictive value of each target cue evolved across trials. Again, this result speaks in favor of a dynamic system that updates the predictive value of cues on a moment-by-moment basis.

Nevertheless, the results from the test stages of Experiments 3 and 4 contradict the predictions made by the associative theory. The Rescorla–Wagner theory assumes that once the predictive values of the target cues have been updated according to the most recent and frequent information, all previous knowledge about the relationships is completely overwritten (the catastrophic forgetting effect). As we have shown, this limitation is not restricted to the Rescorla–Wagner theory. Other associative models, like backpropagation networks, predict the same effect. However, participants' performance did not suffer from catastrophic forgetting. With regard to their diagnostic responses in the test stage, the percentages of correct responses did not depend on whether trial types were initially presented during the first or second half of the learning stage. Moreover, participants were able to respond according to what they had learned during the first half of the learning stage rather than only on the basis of what had subsequently been learned during the second half. Thus, disease 1 was diagnosed more often on *AB* trials than on *B* trials, even though participants had subsequently learned that cue *A* predicted the absence of this outcome. Correspondingly, disease 2 was diagnosed less often on *DE* than on *E* trials, even though cue *D* was subsequently established as a reliable predictor of this outcome. And yet, this is hardly surprising if it is borne in mind that later information altered the predictive value of the target single cues but not of specific configurations of cues. That is, none of the trials presented in Block 2 of the learning stage contradicted the fact that *AB*, taken as a whole configuration of cues, reliably predicted disease 1 or that *DE* predicted the absence of any disease.

Another result that contradicts the theory's prediction is, as described before, that people were able to learn the relationships between the various cues and outcomes arranged in Experiment 4. The task constituted an example of a nonlinear discrimination in the sense that if each of the target cues is coded independently of their accompanying cues, the task becomes unsolvable for the Rescorla–Wagner model, as no set of weights exists that allow the discrimination to be mastered. What, then, is the origin of these erroneous predictions?

The Role of Configural Cues

A plausible solution to all of these problems would be to allow whole configurations of stimuli to act as unique coding units. Let us examine how a configural representational system is compatible with those results from the test stages of Experiments 3 and 4 that were problematic for an elemental theory. The combination of cues *A* and *B* in a single compound (as in *AB* trials) would activate a configural unit (cue *X*) whose predictive value never decreases because no *AB* → no *O* trials are presented. Moreover, when *AC* trials are presented a different configural unit (cue *Y*) would be activated, and no predictive value is gained by this unit, as it is never paired with the outcome. Now, the knowledge about cue *Y* does not catastrophically interfere with what participants know about cue *X*, and thus, disease 1 diagnoses should be expected during the test stage on *AB* trials. Therefore, the changes that occurred to the contingency between cue *A* and the outcome across trials should leave relatively intact knowledge about the predictive value of cue *X*. A similar argument can be made with regard to cue *D* in the other trial order condition. In addition, a configural account can also help us to understand why people did not make more errors on test trials from the first half of the learning phase than on trials from the second half, as later contingencies should not severely affect knowledge about whole configurations. Correspondingly, the task in our Experiment 4 should now be solvable if configural representations of the stimuli are allowed.

Configural cues and associative models. Although there have been different ways of implementing configural representations in specific associative models (e.g., Gluck, 1991; Pearce, 1987, 1994; Rescorla, 1973), only some of them can account for trial order effects on judgments. Specifically, a theory in which compound stimuli have configural representations and in which the constituent elements are also coded (Gluck, 1991; Rescorla, 1973) would leave the effect unexplained. On this theory, the asymptotic associative strengths for cues *A* and *D* would be 0.0 (see below). Alternative configural explanations such as those presented by Pearce (1994) or Kruschke (1992) assume that compounds of stimuli only have direct configural representations and performance is mediated by the degree of similarity between the training and test configurations.

Let us illustrate how Pearce's (1994) configural model can predict both the absence of catastrophic forgetting and the

trial order effect on judgments. The model has been implemented as a connectionist network that includes four layers of units: an input layer, two layers of hidden units (an output layer and a layer of configural units), and a single unit layer that represents the outcome. The activation level of this outcome unit determines the system response. The units in the input layer can be either at an activation level of 0 or 1, depending on whether the element of the stimulus pattern the unit is representing is absent or present, respectively. Each input unit is connected to a single output unit. The intervention of these output units ensures that each stimulus pattern activates maximally (an activation level of 1) a single configural unit. Henceforth, the configural unit can be regarded as representing a particular stimulus pattern. In addition, each stimulus pattern can activate more than one configural unit through a process of generalization. If we assume that configural unit x becomes maximally activated when input pattern X is presented, what will be its activation value when input pattern Y is presented? According to the model, the activation value a_x of configural unit x will be proportional to the degree of similarity between the stimulus patterns. Pearce (1987, 1994) assumes that the similarity of input patterns X and Y is a linear function of the number of elements they share,

$$a_x = n_C\left(\frac{1}{\sqrt{n_X}} \times \frac{1}{\sqrt{n_Y}}\right), \tag{5}$$

where n_C is the number of input units both patterns share and n_X and n_Y are the number of input units that are specific to each stimulus pattern.

Thus, the activation level of the outcome unit when pattern X is presented (V_x) has a double origin. Part of the activation is conveyed by the connection between the configural unit maximally activated and the outcome unit (w_x), and some of its activation comes through the connections between other configural units activated through generalization and the outcome unit,

$$V_x = w_x + \sum_{i=1}^{n}(S_{x,i} \times w_i), \tag{6}$$

where $S_{x,i}$ is the squared activation of these other configural units (see Equation 5) and w_i represents their connections to the outcome unit.

Associative learning involves modifications only in the weight of the connection between the configural unit maximally activated (one for each stimulus pattern) and the outcome unit. These modifications are governed by Equation 7:

$$\Delta w_x = \alpha \times \beta \times (\lambda - V_x). \tag{7}$$

Note the equivalence between Equations 4 and 7. As in the Rescorla–Wagner model, the modification of weights is proportional to an error term, α and β represent learning rates, and λ is set to 1 when the outcome is present and to 0 otherwise.

According to the model, the combination of cues A and B in a single compound will maximally activate a configural unit, which increases its associative strength on $AB \rightarrow O_1$ trials. Moreover, a different configural unit will be activated maximally on $AC \rightarrow$ no O trials, but no associative strength is gained by this unit as it is never paired with the outcome. To illustrate specifically the model's predictions, we have simulated the main results of the experiments. The simulation results were averaged across 10 replications using an equal number of randomly selected sequences of trials. Different sets of simulations were run using different free parameters, but as the results did not substantially vary, we only present simulation data in which the product of the free parameters ($\alpha \times \beta$) was given a value of .5.

Table 9 presents the mean activation values of outcome unit O_1 (in the con–inh trial order condition) and O_2 (in the inh–con condition) for the different trial types presented in the test stages of Experiments 3 and 4 and the activation values of the outcome units when the network was presented the target cues A and D.

The results from the simulation of the test stage of Experiment 3 showed similar proportions of correct diagnoses for those trial types originally presented in Blocks 1 and 2 of the training session. The activations of the outcome units were high for those trial types that were paired with the target outcome and low for those trial types that were paired

Table 9
Simulation Results for Experiments 3 and 4 Using Pearce's (1994) Configural Model

	Experiment 3				Experiment 4			
Trial order condition	Trial types		Target cue		Trial types		Target cues	
Con–inh	*AB*	0.719	*A*	0.134	*AB*	0.893	*A*	0.214
	B	0.031			*B*	0.054		
	AC	0.037			*AC*	0.036		
	C	0.946			*C*	0.948		
Inh–con	*DE*	0.413	*D*	0.436	*DE*	0.098	*D*	0.289
	E	0.977			*E*	0.919		
	DF	0.970			*DF*	0.956		
	F	0.026			*F*	0.033		

Note. The figures represent the mean activation values of outcome units O_1 and O_2 after the training stage in the con–inh and inh–con conditions, respectively. O_1 and O_2 are diseases 1 and 2, respectively; con = contingent; inh = inhibitory. *A*, *B*, *C*, *D*, *E*, and *F* are symptoms.

with the absence of any target disease, regardless of the block in which they were originally presented. With regard to the stimulation of participants' contingency judgments, the activation value of outcome unit O_2 when the target cue D is presented ($M = 0.436$, $SE = 0.002$) is higher than the activation value of outcome unit O_1 when the target cue A is presented ($M = 0.134$, $SE = 0.009$). Thus, the trial order effect found in participants' judgments is predicted by the model.

The results from the simulation of Experiment 4 are very similar. As can be seen, the model can solve the nonlinear discrimination, and the results from the simulation of the test stage do not show catastrophic interference. With regard to the simulation of participants' contingency ratings, the model again predicts a trial order effect, that is, a difference between the activation values of the outcome units O_1 and O_2 when the target cues A and D are presented ($M = 0.214$, $SE = 0.003$ and $M = 0.289$, $SE = 0.003$, respectively). Nevertheless, the trial order effect predicted for Experiment 3 is of a higher magnitude than for this last experiment. Participants' contingency judgments also corroborated this prediction. Specifically, the absolute difference between participants' judgments from the two trial order conditions was 26.8 (45.0 vs. 18.2) in Experiment 3 and 13.2 (52.9 vs. 39.7) in Experiment 4. Correspondingly, the difference in the activation of the outcome units when the target cues are presented is 30.2 (43.6 vs. 13.4, on a comparable scale from 0 to 100) in Experiment 3 and 7.5 (28.9 vs. 21.4) in Experiment 4. The main difference between the model's predictions and the results is the lower magnitude of the trial order effect that the model predicts in Experiment 4. However, this difference should not be taken as a very serious caveat of the model, as it does qualitatively reproduce the general pattern of results.

Figure 5 reproduces the trial-by-trial data from Experiment 2, previously shown in Figure 2, together with the results of a further simulation based on Pearce's model and using the same parameters as before. The model provides an excellent fit to the data.

To conclude, Pearce's configural model allows us to give coherence to the pattern of results we found. The absence of catastrophic forgetting and the influence of trial order in contingency judgments can both be understood in terms of the model. It provides a solution to the catastrophic forgetting problem that involves a trade-off between two empirical constraints, namely, the fact that new information does not completely overwrite prior knowledge and the ability to adapt to the new incoming information through a process of generalization. This particular solution (not the only one—see Lewandowsky, 1994, for further discussion) is based on the creation of semidistributed representations of the incoming information by means of the assignment of exclusive configural representations to each new input pattern and the operation of a generalization mechanism based on pattern similarity.

Because many configural learning phenomena can be explained by a version of the Rescorla–Wagner theory that retains the basic elemental coding scheme for cues but that supplements it with additional input representations for configurations of cues (see Gluck, 1991; Shanks, Charles, Darby, & Azmi, 1997), it is important to note that the present results cannot easily be accommodated by such a model. The clearest problem emerges in Experiment 3. According to this account, the trial types relevant to cue A in that experiment are functionally $ABX \rightarrow O_1$, $B \rightarrow$ no O in Block 1 and $ACY \rightarrow$ no O, $C \rightarrow O_1$ in Block 2, where X and Y are configural cues created by the combinations of A and B, and A and C, respectively. The configural cues are assumed to function just like the other cues in the updating of associative weights. According to the configural-cue version of the Rescorla–Wagner theory, cue A must lose associative strength on the ACY trials of Block 2 and, hence, the proportion of outcome predictions in the test stage to ABX should be lower than at the end of Block 1. Conversely, there should be an increase in the proportion of outcome predictions on DE trials. Table 7 provides some support for these predictions, but also suggests that the effect is not due to the Block 2 trials, because the Block 2 trial types themselves suffer a roughly comparable level of performance decrement in the test stage. This analysis is supported by Shanks, Darby, and Charles (in press) and by Shanks et al. (1997), who included appropriate control groups to show that Block 2 trial types in fact have no detectable impact at all on responding to Block 1 trial types in experimental designs of this sort. Thus it does not seem as if a model that combines elemental and configural representations is adequate; instead, the data require an entirely configural model, such as that developed by Pearce (1994).

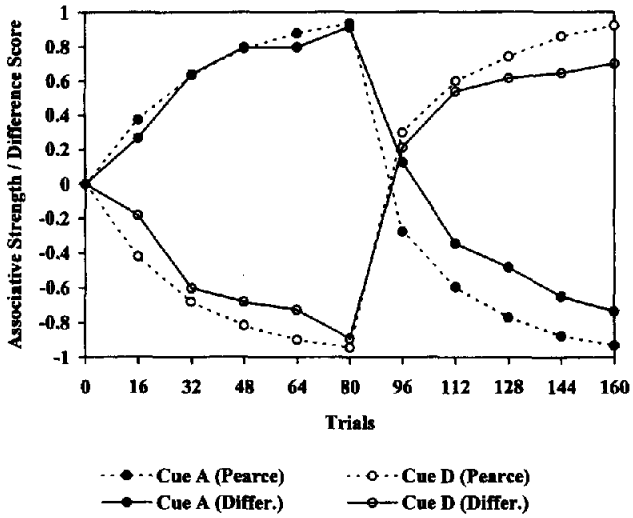

Figure 5. Solid lines: differences (Differ.; cue A) across the learning stage between the proportion of times disease 1 is predicted on AB versus B and AC versus C trials and differences (cue D) between the proportion of times disease 2 is predicted on DE versus E and DF versus F trials in Experiment 2. These are the same data as shown in Figure 2. Dashed lines: predicted changes in the differences of the activation values of the outcome units on AB and B and AC and C trial types (cue A), and on DE and E and DF and F trial types (cue D) across the learning stage, according to Pearce's (1994) theory.

Configural cues and the PCM-ALG. As was the case for associative models, it is possible to conceive of a configural

version of the PCM-ALG. Would such a model be able to deal with our trial order findings? The answer is a clear "no." The weakness of the PCM-ALG is that its basic coding units are conditional probabilities computed across focal sets in which all trial types are evenly represented, and such probabilities are inherently independent of trial order. It does not matter whether participants compute the probability of an outcome given a single symptom, or the probability of an outcome given a configuration of symptoms: In both cases, the resulting probability will be independent of the order of the trials across which it is computed.

Concluding Comments

To sum up, the results reported here clearly show the limitations of the PCM-ALG as an account of how contingency judgments are made. The lack of sensitivity to the distribution of the relevant information across trials represents a critical shortcoming, and this result represents a major limitation of any psychological theory that merely imposes normative or logical constraints on theorizing about psychological mechanisms. Moreover, the results showed the limitations of elemental associative theories such as the Rescorla–Wagner theory, but these results do not invalidate associationist theories as a general framework for understanding how people detect predictive relationships in their changing environment, as Pearce's (1994) configural model was able to provide a good account, within the associationist framework, of all of our main findings. We conclude that processing models based on associationist principles are capable of providing an excellent theory of human contingency judgment.

References

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

Baker, A. G., Berbrier, M. W., & Vallée-Tourangeau, F. (1989). Judgements of a 2 × 2 contingency table: Sequential processing and the learning curve. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 41*(B), 65–97.

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.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105.

Chapman, G. (1991). Trial order affects 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., & Holyoak, K. J. (1995). Complex 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. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58,* 545–567.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40,* 83–120.

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

Curley, S. P., Young, M. J., Kingry, M. J., & Yates, J. F. (1988). Primacy effects in clinical judgments of contingency. *Medical Decision Making, 8,* 216–222.

Davey, G. C. L., & Singh, J. (1988). The Kamin "blocking" effect and electrodermal conditioning in humans. *Journal of Psychophysiology, 2,* 17–25.

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

Fales, E., & Wasserman, E. A. (1992). Causal knowledge: What can psychology teach philosophers? *Journal of Mind and Behavior, 13,* 1–28.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science, 2,* 50–55.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247.

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

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior, 1967: Aversive stimulation* (pp. 9–31). Coral Gables, FL: University of Miami Press.

Kremer, E. F. (1971). Truly random and traditional control procedures in CER conditioning in the rat. *Journal of Comparative and Physiological Psychology, 76,* 441–448.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science, 5,* 3–36.

Lewandowsky, S. (1991). Gradual unlearning and catastrophic interference: A comparison of distributed architectures. In W. E. Hockley & S. Lewandosky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 445–476). Hillsdale, NJ: Erlbaum.

Lewandowsky, S. (1994). On the relation between catastrophic interference and generalization in connectionist networks. *Journal of Biological Systems, 2,* 307–333.

López, F. J., Cobos, P. L., Caño, A., & Shanks, D. R. (in press). The rational analysis of human causal and probability judgment. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition.* Oxford, England: Oxford University Press.

Lovibond, P. F., Siddle, D. A. T., & Bond, N. (1988). Insensitivity to stimulus validity in human Pavlovian conditioning. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 40*(B), 377–410.

Martin, I., & Levey, A. B. (1991). Blocking observed in human eyelid conditioning. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 43*(B), 233–256.

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.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159–188.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in*

parallel distributed processing: A handbook of models, programs, and exercises. Cambridge, MA: MIT Press.

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). New York: Academic Press.

Melz, E. R., Cheng, P. W., Holyoak, K. J., & Waldmann, M. R. (1993). Cue competition in human categorization: Contingency or the Rescorla–Wagner learning rule? Comments on Shanks (1991). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1398–1410.

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.

Price, P. C., & Yates, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition, 21,* 561–572.

Price, P. C., & Yates, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1639–1655.

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

Rescorla, R. A. (1972). Informational variables in Pavlovian conditioning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6, pp. 42–64). New York: Academic Press.

Rescorla, R. A. (1973). Evidence for a "unique stimulus" account of configural conditioning. *Journal of Comparative and Physiological Psychology, 85,* 331–338.

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Shanks, D. R. (1985a). Continuous monitoring of human contingency judgment across trials. *Memory & Cognition, 13,* 158–167.

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

Shanks, D. R. (1987). Acquisition functions in contingency judgment. *Learning and Motivation, 18,* 147–166.

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 433–443.

Shanks, D. R. (1993a). Associative versus contingency accounts of category learning: Reply to Melz, Cheng, Holyoak, and Waldmann (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1411–1423.

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

Shanks, D. R. (1995a). Is human learning rational? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 257–279.

Shanks, D. R. (1995b). *The psychology of associative learning.* Cambridge, England: Cambridge University Press.

Shanks, D. R., Charles, D., Darby, R. J., & Azmi, A. (1997). *Configural and elemental processes in human associative learning.* Manuscript submitted for publication.

Shanks, D. R., Darby, R. J., & Charles, D. (in press). Resistance to interference in human associative learning: Evidence of configural processing. *Journal of Experimental Psychology: Animal Behavior Processes.*

Shanks, D. R., & López, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory & Cognition, 24,* 511–522.

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* (Vol. 34, pp. 265–311). San Diego, CA: Academic Press.

Sloman, S. A., & Rumelhart, D. E. (1992). Reducing interference in distributed memories through episodic gating. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 1, pp. 227–248). Hillsdale, NJ: Erlbaum.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.

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

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

Waldmann, M. R., & Holyoak, K. J. (1997). Determining whether causal order affects cue selection in human contingency learning: Comments on Shanks and Lopez (1996). *Memory & Cognition, 25,* 125–134.

Wasserman, E. A. (1990). Detecting response–outcome relations: Toward an understanding of the causal texture of the environment. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 27–82). San Diego, CA: Academic Press.

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

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

Yates, J. F., & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta Psychologica, 62,* 293–302.

(Appendix follows)

Appendix

Contingency Calculations for the Probabilistic Contrast Model, PCM-ALG

Experiment 1

Cues *B*, *C*, *E*, and *F* are all individually associated to some degree with their respective outcomes and so will be selected as conditionalizing cues. Ideally, participants compute the contrast for a target cue conditional on the absence of all conditionalizing cues, but this is not possible here because *A* and *D* are never seen in isolation. Thus, contrasts must be computed conditionalized on the presence or absence of the known predictors. The contingency ΔP between cue *A* and outcome O_1 conditional on the presence of *B* is 8/10 − 2/10 = .6, and on its absence 8/10 − 8/130 = .74, whereas the contingency conditional on the presence of *C* is 8/10 − 8/10 = 0 and on its absence 8/10 − 2/130 = .78. Correspondingly, the contingency between cue *D* and outcome O_2 conditional on the presence of *E* is 8/10 − 8/10 = 0 and on its absence 8/10 − 2/130 = .78, whereas the contingency conditional on the presence of *F* is 8/10 − 2/10 = .6 and on its absence 8/10 − 8/130 = .74. Thus, the set of contrasts is the same for *A* and *D*.

Experiments 2 and 3

Cues *C* and *E* are the conditionalizing cues. Across the whole experiment, the contingency between cue *A* and outcome O_1 conditional on the presence of *C* is 0 − 1 = −1, whereas the contingency conditional on *C*'s absence is 1 − 0 = 1. Correspondingly, the contingency between cue *D* and outcome O_2 conditional on the presence of *E* is 0 − 1 = −1, whereas the contingency conditional on *E*'s absence is 1 − 0 = 1. Thus, the pair of contrasts is again the same for *A* and *D*.

Across Block 2, however, the $A \rightarrow O_1$ contingency (−1, conditional on the presence of *C*) is greater than the $D \rightarrow O_2$ contingency (1, conditional on the presence of *F*).

Experiment 4

Across the whole learning phase, the contrasts for *A* and *D* are the same as in Experiments 2 and 3. But what if only trials falling within a restricted temporal window are included? Assume that the window only includes trials from the last quarter of the learning phase (the argument is similar for other window sizes). Cues *C* and *E* are again the conditionalizing cues. Using the trial frequencies shown in Table 8, it can be seen that the contingency between cue *A* and outcome O_1 conditional on the presence of *C* is 0/4 − 4/4 = −1, calculated across the *AC* and *C* trial types, whereas the contingency conditional on *C*'s absence is

$$\Delta P = P(O_1/A \cdot \overline{C}) - P(O_1/\overline{A} \cdot \overline{C}) = 1/1 - 0 = 1,$$

calculated across the remaining trials. For *D*, the corresponding contrasts are 0/1 − 1/1 = −1 conditional on the presence of *E* and 4/4 − 0 = 1 conditional on its absence. Again, the pair of contrasts is the same for *A* and *D*.

Cheng and Holyoak (1995, p. 286) proposed that in some circumstances participants may rely on unconditional contingencies, and the unconditional ΔP values are indeed different for cues *A* and *D* in the present experiment. For example, across the last block of trials the unconditional ΔP is 1/5 − 4/15 = $-0.0\overline{6}$ for *A* and 4/5 − 1/15 = $0.7\overline{3}$ for *D*, and the direction of this difference is consistent with participants' judgments. However, the circumstances specified by Cheng and Holyoak for the use of unconditional contingencies do not apply in the present experiment. According to their theory, unconditional contingencies are incorporated in the judgment process only if some of the contrasts conditional on the presence and absence of the conditionalizing cues cannot be computed, or if ceiling effects are present. In the present experiment, as *C* is the only conditionalizing cue for *A* and *A*'s contingency conditional on both the presence and absence of *C* can be computed (and similarly for *D*), there is no need for the use of unconditional contingencies. Also, there are no ceiling effects. Thus, for participants to rely on unconditional contingency would, according to the PCM, be highly nonnormative in the present circumstances, and the theory predicts that participants should not do so. Further empirical evidence against the idea that the judgment process combines conditional and unconditional contrasts has been presented by Shanks (1993a).

Received May 14, 1996
Revision received August 27, 1997
Accepted September 2, 1997 ■