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.

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. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin, 113,* 211–228.

Wasserman, E. A., Dorner, W. W., & Kao, S. -F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 509–521.

Williams, D. A., Sagness, K. E., & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 694–709.

Williams, D. A., Travis, G. M., & Overmier, J. B. (1986). Within-compound associations modulate the relative effectiveness of differential and Pavlovian conditioned inhibition procedures. *Journal of Experimental Psychology: Animal Behavior Processes, 12,* 351–362.

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

Young, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin & Review, 2,* 83–104.

DISTINGUISHING ASSOCIATIVE AND PROBABILISTIC CONTRAST THEORIES OF HUMAN CONTINGENCY JUDGMENT

David R. Shanks
Francisco J. Lopez
Richard J. Darby
Anthony Dickinson

I. Introduction

In an article published in 1984, Dickinson, Shanks, and Evenden suggested and provided some supporting evidence for the idea that human judgments of event contingency might be profitably analyzed from an associationist perspective, and more particularly that they might be understood in terms of the processes embodied in prevailing associationist theories of animal conditioning. With the benefit of hindsight, an associative perspective on contingency judgment does not seem especially radical. In a contingency judgment task, subjects are asked to rate the extent to which an action and outcome or a cue and outcome are related, and the parallels to (respectively) instrumental and Pavlovian conditioning procedures should be obvious.

In the years that have followed, the associative view of contingency judgment has been explored in a large number of studies and has generated a good deal of controversy (as the existence of this volume demonstrates). In this chapter, we survey some of the main aspects of this field, and in particular we review some of the progress that has occurred since our earlier evaluation of associative theory (Shanks & Dickinson, 1987) in this

THE PSYCHOLOGY OF LEARNING AND MOTIVATION, VOL. 34

Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

series. In that review, we presented a number of experiments that were consistent with an associationist account of contingency judgment, but we also noted several deeply problematic findings. We concentrate in the present chapter on reevaluating those problematic results in the light of more recent investigations.

Why has the study of contingency judgment expanded into such a large area of research interest? Doubtless there are very many reasons, but it is important to recognize three of the main ones. First, a moment's reflection reveals that many aspects of our daily lives require us to make accurate evaluations of event contingencies. In general, we perform an action because we believe that an outcome that we value or wish to avoid covaries with that action: we take exercise, for instance, because we believe that it covaries with good health, and we stop smoking because it causes lung cancer. Similarly, many of our judgments depend on accurate assessments of event covariations: a doctor makes a particular disease diagnosis because he or she has learned that certain symptoms covary with the disease. Thus, it is not hard to recognize the potential importance of laboratory studies of the mechanisms of contingency judgment.

Second, the idea that human judgments and animal conditioning may be served by similar mechanisms appears to have evoked strong reactions in many researchers, with some finding the idea highly appealing and others totally unacceptable. Those who find the idea appealing are likely to be enthusiastic about the increasing impact of associationism and connectionist theories in cognitive psychology, which has led many researchers to look more closely at the basic mechanisms of learning. Although a widespread view in psychology is that the best place to study such mechanisms is in the animal laboratory, it has been increasingly recognized that certain issues concerning elementary learning processes are more amenable for study with humans than with animals. Contingency judgment tasks represent one example in which such processes can be readily investigated (another would be studies of category learning; see Nosofsky, 1992; Shanks, 1994).

In contrast, those researchers who are unhappy with the idea that similar processes may underlie animal conditioning and human contingency judgment tend to point to the latter's more "high-level" and "cognitive" aspects, which they contrast with the relative automaticity and cognitive impenetrability of conditioning. For example, Waldmann and Holyoak (1992; see Waldmann, this volume, Ch. 2) emphasized the way in which subjects might bring complex causal theories to bear in making contingency judgments. Although we would dispute the claim that contingency judgments are always mediated by complex and high-level cognitive processes (see Section III), we hope that the present chapter will make it clear that we do not regard associative mechanisms as "dumb." On the contrary, contemporary animal learning theories have to appeal to quite rich representational structures to explain even such simple behaviors as approach to a food source (see Dickinson, 1980; Dickinson & Shanks, 1995).

The third reason why contingency judgment has become the focus of so much research interest is that it provides a very simple domain in which questions about rationality can be addressed. Dickinson et al.'s (1984) paper came against a backdrop of studies concerned not so much with explaining contingency judgment but rather with determining whether humans are capable of normatively accurate judgments (e.g., Allan & Jenkins, 1980; Jenkins & Ward, 1965; Smedslund, 1963; Wasserman, Chatlosh, & Neunaber, 1983). Like the simple decision making tasks studied by Kahneman and Tversky (1972), contingency judgment experiments seemed to be easy to analyze from a normative point of view and also seemed to some to yield clear evidence of violations of normative behavior. Although we will only briefly consider this issue in the present chapter, it is important to note that the associative view of contingency judgment has a good deal to say about normativeness (see Shanks, 1995a, 1995b).

The plan of the present chapter is as follows. We begin by briefly describing the well-known conditioning theory of Rescorla and Wagner (1972), which serves as our associationist starting point, and we contrast it with a new nonassociationist theory that has recently generated a good deal of interest (see Chapters 8 & 2, respectively, by Cheng, Park, Yarlas, & Holyoak and by Waldmann, this volume), namely Cheng and Holyoak's (1995) "probabilistic contrast" theory. We then consider three problems that seem to be particularly acute for the associationist approach (the apparent lack of convergence of judgments under different noncontingent schedules, the influence of causal models, and the existence of a phenomenon known as retrospective revaluation). We describe new results that suggest that these problems may not after all be too damaging to associationist theories. Instead, we conclude that it is the probabilistic contrast theory that has the greatest difficulty accounting for our results. Finally, we report some new results that suggest that the basic units of analysis in associationist theories, namely independent "elemental" representations of stimuli, are inadequate. We consider ways in which "configural" representations might be incorporated into associative theories of contingency judgment.

A. The Rescorla–Wagner Theory

In an influential animal conditioning experiment, Saavedra (reported by Wagner, 1969) demonstrated that the extent to which a cue and an outcome come to be associated is a function not simply of their repeated pairing but rather of the degree to which the cue is a reliable predictor of the

outcome. Saavedra's experiment used a rabbit eyelid conditioning procedure. In this procedure, a brief neutral stimulus such as a light is paired with a mild shock to the eye. As a result, it becomes a conditioned stimulus (CS) capable of eliciting by itself the conditioned response of blinking. In one group (called the *contingent* group) the light was paired on some trials with another neutral stimulus, a tone, and this compound was followed by shock. On other trials the tone occurred on its own and was unreinforced. If the light is cue A and the tone cue B, and shock is the outcome (O), then the animals in this group received intermixed AB → O and B → no O trials, with shock being contingent on cue A's presence. At the end of the training phase, cue A was tested on its own, and (unsurprisingly) elicited a substantial conditioned response, indicating that it had become strongly associated with shock.

A second group of animals (the *noncontingent* group) again received AB → O trials, but for these animals trials with cue B by itself were reinforced rather than unreinforced; thus these animals received intermixed AB → O and B → O trials. Since shock could occur on trials with B alone, it is not contingent on cue A's presence. When this group was tested with cue A by itself, little conditioned responding was elicited, indicating that cue A had become only weakly (if at all) associated with the shock.

The importance of Saavedra's result (and of a number of other, similar results obtained around the same time; see Mackintosh, 1983) lies in the fact that even though cue A was paired with the outcome an equal number of times in the two groups, its association with the outcome differed. The reason is that in the contingent group, A was a good predictor of the outcome compared to cue B, whereas in the noncontingent group, A was an entirely redundant stimulus in that it provided no information that was not already conveyed by B. Thus, a process of *cue selection* appears to operate in associative learning: a cue will be selected for association with another event only if it is a good relative predictor of that event.

This sort of cue selection effect played a major role in the late 1960s in the development of formal associationist learning theories such as that of Rescorla and Wagner (1972), which is equivalent to the "delta" rule used to update the weights in many current connectionist models of human learning (e.g., Gluck & Bower, 1988; Kruschke, 1992, 1993; McClelland & Rumelhart, 1985). The Rescorla–Wagner theory is just one of a number of accounts of associative learning that were developed in response to demonstrations of selective conditioning. In one way or another, however, they all deploy the idea that learning is controlled by the relative predictive validity of a cue as assessed by an expectancy error term. An error exists whenever there is a mismatch between what is predicted to occur on a trial and what actually occurs. Attentional theories, for example, argue that the

absolute (Pearce & Hall, 1980) or relative (Mackintosh, 1975) error term generated on a learning episode determines the subsequent associability of a cue with the outcome. Although the relative merits of these various theories is still a matter of dispute within the context of conditioning, we shall focus on the Rescorla–Wagner model in the present discussion. Rescorla and Wagner's (1972) theory explains Saavedra's result by assuming that there is a ceiling or limit to the amount of association strength that can be supported by any given outcome event. On each trial, the strength (V) of the target association is changed by amount dV, proportional to the error ($\lambda - \Sigma V$):

$$dV = \alpha\beta\,(\lambda - \Sigma V), \tag{1}$$

where α is a learning rate parameter determined by the salience of the cue, β is another learning rate parameter determined by the salience of the outcome, λ is the associative strength that is required to predict fully the occurrence of the outcome, and ΣV is the sum of the associative strengths of all cues present on the current trial. It is assumed that a subject's judgment about the relationship between a cue and the outcome is monotonically related to the associative strength of that cue (V).

How does the model generate the effect observed by Saavedra? Learning will cease when the outcome is fully predicted on each trial. In the noncontingent condition, involving AB → O and B → O trials, cue B must have associative strength of λ in order that the outcome is predicted on the B → O trials. For the outcome also to be fully predicted on the AB → O trials, A must therefore have associative strength of zero. In the contingent condition, by contrast, the B → no O trials mean that B's asymptotic strength is zero. This in turn requires that A have associative strength of λ in order that the outcome is fully predicted on the AB → O trials. Hence A has associative strength of zero in the noncontingent condition and λ in the contingent one.

Another important cue selection effect is the well-known phenomenon of "blocking." This phenomenon, originally reported in animal conditioning by Kamin (1968), refers to the fact that when a compound of two cues is paired with an outcome, the amount learned about one of the cues (the target) is reduced if the other (competing) cue has been separately pretrained as a predictor of the outcome. Under these circumstances, the competing cue is said to "block" learning about the target cue. As with Saavedra's results, the importance of blocking lies in its illustration that simple associative learning is informationally sensitive and reflects the relative predictive validity of a cue. Although the target cue is presented in temporal contiguity with the outcome—a condition traditionally thought

to be sufficient for learning – the target is in fact informationally redundant in the sense that the occurrence of the outcome is fully predicted by the pretrained, competing cue.

From the point of view of human contingency judgment, cue selection effects are critical because if they can be demonstrated to occur in humans, that would encourage the claim that an associationist mechanism is responsible for generating such judgments. The idea would be that judgments are based on the strength of a mental bond or association connecting representations of the predictive event and outcome. And, indeed, it turns out that cue selection effects are easy to obtain with human subjects (see Spellman, this volume, Ch. 5). For instance, a result comparable to that obtained by Saavedra has been observed in humans. Shanks (1991) presented subjects with hypothetical medical patients who had different symptoms, and subjects had to predict which disease each patient had. When the pattern of symptom–disease pairings was analogous to Saavedra's contingent condition, in that subjects saw AB → O and B → no O trials, ratings of the contingency between symptom A and the disease were high. But when the trial types conformed to the noncontingent arrangement (AB → O, B → O), ratings were much lower. Cue-selection effects have now been observed in human contingency judgment by a number of researchers (e.g., Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Chapman, 1991; Chapman & Robbins, 1990; Price & Yates, 1993; Shanks, 1985b, 1989, 1991; Wasserman, Elek, Chatlosh, & Baker, 1993; Williams, Sagness, & McPhee, 1994; see Chapters 1, 6, 5, & 3, respectively, by Baker, Murphy, & Vallée-Tourangeau; Wasserman, Kao, Van Hamme, Katagiri, & Young; Spellman; Williams, this volume). The existence of such effects is highly encouraging for the notion that contingency judgment can be interpreted within an associationist framework of the sort provided by Rescorla and Wagner (1972).

B. The Probabilistic Contrast Model

In a challenge to this associationist approach, Cheng, Holyoak, and their colleagues have argued that cue-selection effects emerge for entirely different reasons: subjects do not increment and decrement mental associations, but instead base their judgments on the difference between the conditional probability of the outcome given the presence versus the absence of the target predictor. A large and very interesting research program is being pursued by these authors and their colleagues that challenges the enterprise of understanding human and animal learning in terms of associationist principles (Cheng, 1993; Cheng & Holyoak, 1995; Cheng & Novick, 1990, 1991, 1992; Holyoak, Koh, & Nisbett, 1989; Melz, Cheng, Holyoak, & Waldmann, 1993; Waldmann & Holyoak, 1992). In fact, it is fair to say that the emergence of this theory has been the main theoretical development of the last decade in this field.[1]

So-called "contingency" theories take as their starting point the idea that judgments of contingency are computed via a mental version of the normative Δp equation. Subjects maintain mental records of the conditional probability of the outcome given the target cue, $p(O/C)$, and of the conditional probability of the outcome in the absence of the target cue, $p(O/-C)$, and base their judgments on the difference between these probabilities:

$$\Delta p = p(O/C) - p(O/-C). \tag{2}$$

The probabilistic contrast model (PCM) constitutes a normative generalization of the standard Δp measure to situations in which multiple predictors or cues are involved and where the background for a given target cue may not be constant. In these more complex situations, it is not just the relationship between the target cue (A) and the outcome that is taken into account but also the predictive status of other cues. Specifically, the contingency between cue A and the outcome must be computed over certain restricted *focal sets* of events. Suppose that A co-occurs with some background cue that we will designate B. Since the true predictive value may lie with B rather than with A, one focal set is formed by the set of events in which cue B is always present. Then, the contingency between A and the outcome conditional on the presence of this alternative cue is just 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:

$$\Delta p = p(O/A.B) - p(O/-A.B) \tag{3}$$

Thus, Equation 3 represents a special case of Equation 2 in which the trial types over which contingency is computed are restricted to those in which the alternative cue B is present, and it should be fairly straightforward to see that Equation 3 represents a "contrast" between what happens when A is present versus when it is absent, holding B's presence constant. In

[1] It should be noted that the probabilistic contrast model 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 chapter, we are solely concerned with the algorithmic-level version of the PCM formulated by Cheng and Holyoak (1995) and Melz et al. (1993) and directly contrasted by them with the Rescorla–Wagner theory.

accordance with this simple idea, the values of Δp the equation yields for the contingent and noncontingent conditions of Saavedra's experiment differ in the right direction. In both the contingent ($AB \rightarrow O$, $B \rightarrow$ no O) and noncontingent ($AB \rightarrow O$, $B \rightarrow O$) conditions, cue B is the alternative predictor and the focal set contains all trials on which B is present. Applying Equation 3, it is easy to see that $\Delta p_A = 1$ in the contingent condition and 0 in the noncontingent one.

In fact, many different contrasts can potentially be computed that have the same form as Equation 3. If the target cue covaries with many other cues, then multiple contrasts can be calculated in which different background cues are held constant. Thus, in a design involving $AB \rightarrow O$, $B \rightarrow$ no O, $AC \rightarrow$ no O, and $C \rightarrow O$ trials, it is possible to compute one contrast for the focal set of trials in which B is present and another one for the focal set of trials in which C is present. These contrasts yield values of $\Delta p_A = 1$ in the former case and $\Delta p_A = -1$ in the latter, reflecting the fact that A predicts the outcome when it is paired with B and predicts its omission when paired with C. The PCM assumes that when multiple contrasts can be computed and the subjects have to provide a single contingency judgment, some integration process is employed to collapse the outputs of the contrasts. For our purposes we will simply assume that this integration process is monotonic such that if all of the contrasts for one cue are greater than the comparable contrasts for a second cue, then judgments will be greater for the first than for the second cue.

Cheng and Holyoak's (1995) implementation of the PCM includes several steps: (1) selecting which cues will be incorporated in the focal sets; (2) choosing the conditional contrasts to be calculated; and (3) determining how the information provided by the different conditional contrasts is integrated within a single judgment about the predictive value of the target cue. To see how this implementation works, let us consider a specific application of the model to another experiment by Shanks (1991) involving a slightly different version of Saavedra's design in which subjects saw $AB \rightarrow O_1$, $B \rightarrow$ no O, $C \rightarrow O_1$, $DE \rightarrow O_2$, $E \rightarrow O_2$, and $F \rightarrow$ no O trials, where the letters are medical symptoms and O_1 and O_2 are different diseases. With regards to the relationship between the target cue A and outcome O_1, the implementation of the model specifies that subjects will initially select cue C as a conditionalizing cue given that this cue has been consistently paired on its own with the outcome, the outcome never occurs outside the experimental trials, and subjects have no prior beliefs about the predictive values of the different cues. However, because A and C are never paired, the only meaningful contrast that can be computed for A is the contrast conditional on the absence rather than presence of cue C. According to this contrast, subjects will compute the probability of O_1 given the presence of A and the absence of C, $P(O_1/A.\ C)$, which has a value of 1. In addition, subjects will compute the probability of O_1 in the absence of both A and C $[P(O_1/-A.-C)]$, which has a value of 0. Thus, cue A has a conditional contingency of $\Delta p = 1$ in the focal set in which cue C is absent.

With regards to the relationship between the target cue D and disease 2, cue E will be selected as a conditionalizing cue, as it has been consistently paired with disease 2. In this case the only conditional contrast that can be computed is one conditional on the presence of cue E (in fact, the target cue D never occurs in the absence of this conditionalizing cue). Thus, subjects will compute the probability of O_2 given the presence of the target cue D and cue E $[P(O_2/D.E)]$, which is 1, and the probability of this disease given the absence of cue D and the presence of cue E $[P(O_2/-D.E)]$, which is also 1. Therefore, cue D has a conditional contingency of $\Delta p = 0$ in the focal set in which cue E is present. Hence, it is straightforward to generate higher ratings for A than for D from the contrasts, which was the outcome observed in the subjects' ratings.

The probabilistic contrast model represents a major advance in our conception of normative theories of contingency judgment because it is readily able to explain a broad range of cue selection effects. The key idea behind the PCM is simply that the evaluation 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, and this idea should require little justification in the context of a scientific methodology that emphasizes the use of controlled experiments that adopt exactly this procedure: in scientific experiments, the researcher sets up an experiment and a control condition in which everything is held constant other than the presence versus absence of the critical factor. The crucial question for our present purposes, of course, is whether we can discriminate between associative and contingency-based theories.

II. Convergence in Noncontingent Conditions

Plainly, the Rescorla–Wagner and probabilistic contrast models stand as powerful alternative descriptions of the contingency judgment mechanism, and each is able to explain the basic fact that variations in objective contingency influence subjects' judgments. Thus, more subtle methods must be developed for pitting the theories against one another, and in the present section, we consider a phenomenon that we identified in our earlier chapter (Shanks & Dickinson, 1987)—concerning contingency judgments under noncontingent conditions—which seems to present a major problem for the Rescorla–Wagner theory. Suppose that subjects are required to make

contingency judgments after varying numbers of trials. Briefly, the theory predicts that when the objective contingency between a target cue or action and the outcome is zero, judgments should converge to an asymptotic value of zero (see Melz et al., 1993; Shanks, 1993). The reason for this is plain: as we saw for the noncontingent (AB → O, B → O) condition in Saavedra's study, A's asymptotic associative strength must be zero in order for the outcome to be accurately predicted on both trial types. Note that the prediction is the same even if $p(O/A)$ and $p(O/-A)$ are less than 1.

The problem with the theory is that, in contrast to this prediction, judgments in a number of experiments seem to be strongly affected (even at asymptote) by the overall probability of the outcome (Baker, Berbrier, & Vallée-Tourangeau, 1989; Shanks, 1985a, 1987). Moreover, in some of these experiments (e.g., Shanks, 1987) judgments were actually negative when the probability of the outcome was low (i.e., .25). Although such findings are at variance with the predictions of associationist theories, they are quite easily explained by contingency theory because it is reasonable to assume that subjects give different weights to the terms in the contingency equation. Evidence from a different domain (see Kao & Wasserman, 1993; Wasserman, Dorner, & Kao, 1990; Wasserman et al., this volume, Ch. 6) has shown that subjects give more weight to trials on which the target cue occurs than to trials on which it is absent. In terms of the Δp formula, this simply means adding weighting parameters a and b ($a > b$) such that:

$$\Delta p = a \cdot p(O/C) - b \cdot p(O/-C). \quad (4)$$

The important cases for our present purposes are noncontingent conditions in which $p(O/C) = p(O/-C) = p(O)$. The equation then becomes:

$$\Delta p = p(O) \cdot (a - b).$$

With different values of $p(O)$, the obtained values of Δp from this equation will converge to a common asymptote only when $a = b$. Whenever $a > b$, Δp (and hence judgments) will be an increasing function of the probability of the outcome, $p(O)$, and hence will not converge to a common asymptote. This, of course, is exactly what was observed in the experiments mentioned above (e.g., Shanks, 1987). Thus, with the simple addition of some well-motivated psychological weighting parameters, a contingency-based theory can readily account for the observed absence of convergence.

Despite the apparent difficulty the empirical results present for associative theories, it is possible that some undesirable experimental artifact is responsible for the convergence results. What might such an artifact be? The likeliest candidate is the use of a within-subjects design in all of the

experiments referred to previously, which creates the possibility that judgments from one condition to another are not independent. Although Shanks (1985a) failed to support this suggestion using a rather indirect method, we decided in our first experiment to look at learning curves in a completely between-subjects design.

A. Experiment 1

In this study, subjects made judgments of contingency across 40 trials in an experiment that used a task developed by Baker et al. (1989). On each trial a tank moved across the computer screen through a minefield and was either blown up by a mine or not. On half the trials the tanks were camouflaged and were able to avoid the mines, which were color sensitive. Every five trials, subjects judged the relationship between the cue (camouflage) and the outcome (avoiding being blown up), using a scale from −100 to +100 (they also gave a confidence rating for each judgment, but these data are not presented here). In addition, once subjects had seen the status of the tank on a given trial they had to predict whether the tank was going to explode or not on that particular trial.

There were four groups in the experiment ($n = 20$ per group), each seeing a different combination of $p(O/C)$ and $p(O/-C)$. For one group, there was a positive but imperfect contingency of $\Delta p = .50$, resulting from values of $p(O/C)$ and $p(O/-C)$ of .75 and .25, respectively. In another condition, the contingency was −.50, derived from values of $p(O/C) = .25$ and $p(O/-C) = .75$. Finally, in two conditions Δp was zero. This was achieved in one case by having two quite high conditional probabilities [$p(O/C) = p(O/-C) = .75$] and in the other case by having two low probabilities [$p(O/C) = p(O/-C) = .25$]. For the positive contingency condition, the Rescorla–Wagner theory predicts that as more and more trials are presented, an increasing but negatively accelerated learning function should appear. The reason for this is that on early trials, the discrepancy between the asymptote of learning (λ) and the combined associative strength of the cues present (ΣV) will be large and hence the increment to associative strength will be large. However, as learning proceeds, this discrepancy will get smaller and smaller (and likewise for the strength changes) until learning ceases when $\lambda = \Sigma V$. In the negative contingency condition, where the outcome is less likely in the presence than in the absence of the cue, judgments should become negative and take a form that is roughly the mirror image of the curve for positive contingencies; that is to say, judgments should drop rapidly in the first few trials and then fall more slowly before reaching a negative plateau.

As can be seen from Fig. 1, judgments started close to zero in each condition. When there was a positive contingency of $\Delta p = .50$, judgments

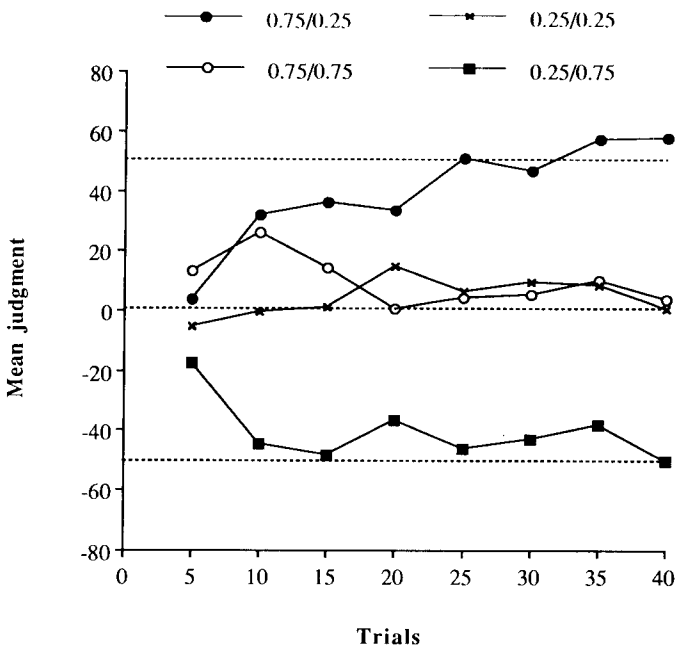

Fig. 1. Mean judgments of contingency across 40 trials under four different cue–outcome contingencies. Judgments were made on a rating scale from −100 to +100. Each condition is designated by two numbers, the first being p(O/C) and the second p(O/−C). Judgments increase under the positive (.75/.25) contingency and decrease under the negative (.25/.75) contingency, in each case yielding terminal judgments close to the actual contingencies (× 100). In the noncontingent conditions (.75/.75 and .25/.25), judgments converge to zero, but when the probability of the outcome is high (in the .75/.75 condition), early judgments are erroneously positive.

increased steadily across trials toward an asymptote of around 50, and when the contingency was −.50, judgments decreased across trials. Crucially, in the noncontingent conditions judgments stayed close to zero across trials, although in the .75/.75 case there was a preasymptotic increase and then decrease during the early trials. This latter effect is important, because it replicates previous observations (e.g., Shanks, 1987) and is predicted by the Rescorla–Wagner theory, as we show later. There is no suggestion that judgments differed at asymptote in the two noncontingent conditions, and accordingly we suggest that previous failures to obtain convergence have been biased by the use of within-subjects methodologies. Of course, this conclusion would be bolstered by a single experiment in which some subjects saw a single condition and others saw all four conditions, but we feel that

the present results are sufficient to reestablish confidence in the predictions of the Rescorla–Wagner theory.

B. Simulations of Experiment 1

In a series of simulations we tried to see whether the Rescorla–Wagner theory and PCM can accurately predict the pattern of results obtained in the different groups of Experiment 1. Intuitively, it would seem unlikely that a normative theory such as the PCM could, in principle, predict the changes of judgments seen across trials as the conditional probabilities on which contrasts are computed remained constant across trials. In our experiment, each block of eight trials contained four trials with the cue and four without. Of the four trials of each type, exactly one or three were paired with the outcome, depending on whether the relevant conditional probability was .25 or .75. Thus, for each block of eight trials, Δp was *exactly* −.5, 0, or .5, and remained constant across trials for each eight-trial block. But while Δp was constant, judgments were not.

Although the fact that judgments in the noncontingent conditions converged suggests that differential weighting of p(O/C) and p(O/−C) is unnecessary, in our PCM simulations we included the two weighting parameters a and b, as specified in Equation 4. We computed the contingency judgments derived from the PCM, taking (weighted) estimations of the conditional probabilities. The predicted judgment on any given trial was simply the difference between the weighted conditional probabilities estimated up to that moment (× 100). The mean predicted judgments were compared with subjects' mean actual judgments and a sum of squared errors (SSE) was calculated between mean observed and predicted judgments across all contingency groups (8 × 4 data points). A search was made to find the values of the two free parameters a and b in the [0,1] interval, which provided the best fit to the actual judgments (i.e., which minimized SSE). The best fitting solution was obtained when a = .93 and b = .85. The parameters suggest that slightly different weights were given to p(O/C) and p(O/−C), in accordance with previous results. As the actual sequence of events that each subject experienced was recorded during the experiment, the theory's predictions could be computed for trial sequences identical to those presented to the subjects. The results are therefore averaged across 20 simulated subjects receiving the same sequences of events as the actual subjects. Figure 2 shows the judgments the PCM predicted under these circumstances.

Plainly, at a qualitative level the fit is not very good in that unlike the actual judgments, the predicted judgments hardly change across trials. Moreover, the model fails to predict any preasymptotic difference in ratings

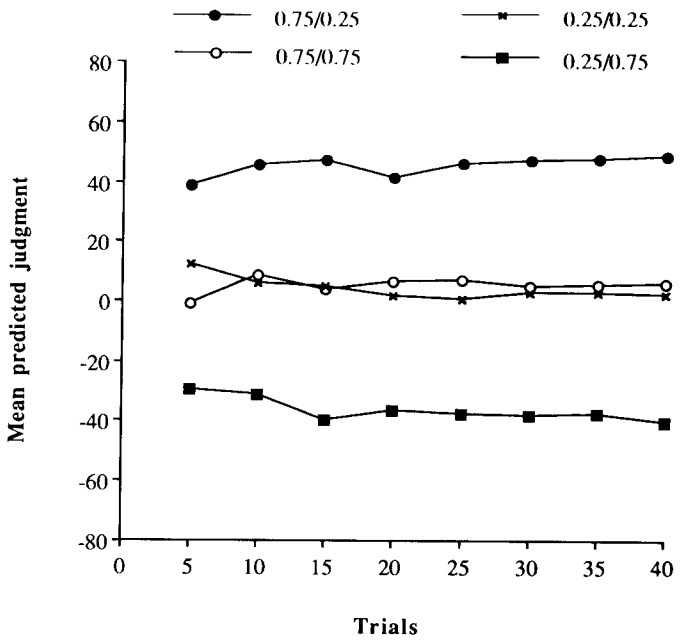

Fig. 2. Best fitting PCM predictions for the data shown in Fig. 1. The figure shows mean computed contingencies across 40 trials under 4 different contingencies: positive (.75/.25), zero (.75/.75 and .25/.25), and negative (.25/.75), where the first figure refers to $p(O/C)$ and the second to $p(O/-C)$.

for the two noncontingent conditions. Can the model be easily revised in order to improve the predictions? One possibility is that subjects start each problem with the assumption that the contingency is zero, and their estimates are then updated by a Bayesian process that incorporates such prior beliefs (see Fales & Wasserman, 1992). But although such an account would allow judgments to increase and decrease across trials under positive and negative contingencies, there is no reason to suppose that preasymptotic divergence in the two noncontingent conditions would be generated. Clearly, further explorations are needed to see if contingency theories can reproduce acquisition data.

In contrast to this poor fit, the Rescorla–Wagner theory provides a very good account of the data. Again, the results are based on 20 simulated subjects receiving trial sequences identical to those presented to the subjects. Four free parameters were included and these were constrained within the [0,1] interval. In this case one pair of parameters was for the salience (α) of the target cue (camouflage) and the background contextual cues,

and the other pair was for the salience (β) of the occurrence and nonoccurrence of the outcome. The Rescorla–Wagner theory assumes that outcomes occurring in the absence of the target cue become associated with contextual cues that compete for the limited associative strength supportable by the outcome. λ was given a value of 100 in order to scale the associative strengths into the same range as subjects' judgments. Again, the mean values of the predicted judgments were calculated and compared with subjects' mean judgments. The parameters chosen were .7 and .3 for the salience (α) of the camouflage and the contextual cues, respectively, and .5 and .9 for the salience (β) of the outcome and of no outcome, respectively. Figure 3 shows the predicted judgments.

As can be seen, the associative model provides a much better fit to the results than the PCM. First, clear changes in judgments are predicted across trials. Second, the preasymptotic divergence of judgments in the noncontingent conditions is reproduced (see Rescorla, 1972; Shanks, 1985a, 1995a, for similar predictions). In the .75/.75 condition, the associative strength of

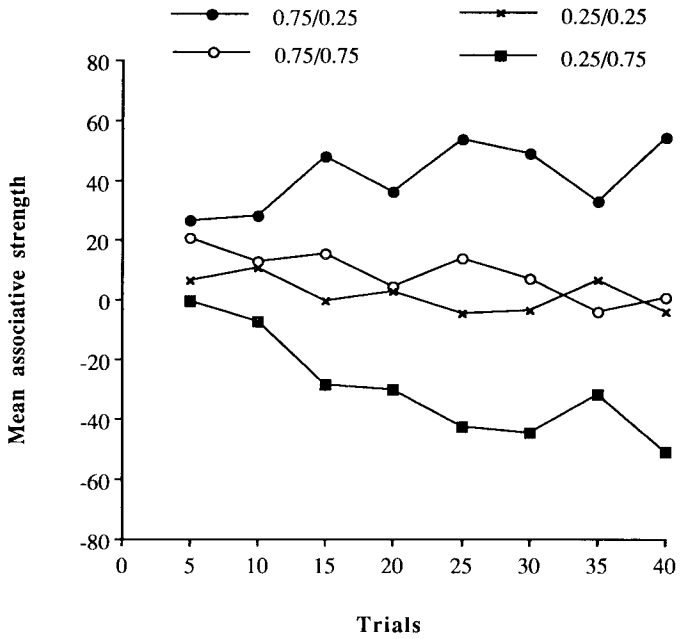

Fig. 3. Rescorla–Wagner theory predictions for the data shown in Fig. 1. The figure shows mean associative strengths across 40 trials under 4 different contingencies: positive (.75/.25), zero (.75/.75 and .25/.25), and negative (.25/.75), where the first figure refers to $p(O/C)$ and the second to $p(O/-C)$.

the target cue becomes substantially positive on early trials, just as in the subjects' judgments. In sum, the results from this experiment are much more conducive to an associative than to a contingency-based explanation. Not only did judgments converge under noncontingent conditions, but the shapes of the learning curves were much better fit by the Rescorla–Wagner theory than by the PCM.

III. Causal Models and Causal Order

Another recent challenge to the associationist view of contingency judgment has come from some studies by Waldmann and Holyoak (1992) on the role of background causal knowledge in subjects' judgments (see also Miller & Matute, this volume, Ch. 4; Matute, Arcediano, & Miller, 1996; Waldmann, this volume, Ch. 2). In contrast to the sort of mechanism embodied in the Rescorla–Wagner theory, Waldmann and Holyoak adopt a mentalistic "causal model" approach to learning, in which "people use meaningful world knowledge, often of a highly abstract sort, to guide their learning about new domains. One major example of abstract world knowledge is knowledge about the basic characteristics of causal relations, such as the temporal precedence of causes to their effects" (Waldmann & Holyoak, 1992, p. 224). According to this account, subjects use real-world knowledge to interpret the cues and outcomes in an experiment, and in particular use knowledge about such things as causal precedence: regardless of the order in which the events are perceived, subjects know about the directionality of causation.

Having brought an interpretative causal mental model to bear on the task, subjects then determine (via Equation 3) the extent to which the cause and effect are related, with the cue being the cause and the outcome the effect. Waldmann and Holyoak noted that different predictions can be derived from the associationist and causal model theories in so-called "diagnostic" learning tasks. Suppose the subject is required to learn some cue–outcome relationship, but the appropriate causal interpretation is that the event labeled as the outcome would (in the real world) have caused the event labeled as the cue rather than vice versa. For instance, in a number of studies (e.g., Shanks, 1991), subjects have learned to predict diseases on the basis of symptoms, but, of course, in the real world it is diseases that cause symptoms rather than vice versa. This sort of "diagnostic" learning can be distinguished from "predictive" learning where the cues are the causes and the outcomes the effects.

Waldmann and Holyoak predicted that while cue selection may emerge in predictive contexts, it should not emerge in diagnostic conditions. In the

noncontingent AB → O, B → O design, for example, the mental model theory assumes that the subject interprets these trial types as $E_1E_2 \leftarrow$ cause and $E_2 \leftarrow$ cause, where E_1 and E_2 are different effects. Despite the fact that the order in which the events are experienced is opposed to their true causal order, on this theory subjects will learn relationships from the cause to the effects, and the presence of E_2 should not alter in any way the judged causal relationship between the cause and E_1. This is because the presence of multiple effects does not change the computed value of Δp for a given cause–effect relationship. As Waldmann and Holyoak say (1992, p. 224), "People have a strong disposition to learn directed links from causes to their effects, rather than vice versa, even in situations in which they receive effect information prior to cause information," and (p. 226) "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."

For an associative account, however, the real-world interpretation of the events is immaterial. If the subject is required to predict outcomes from cues, then the cues represent the input to the system and the outcome represents the target to be predicted, regardless of causal order. Thus, even though cue selection in predictive situations (where the cues are interpreted as causes and the outcomes as effects) can be explained by either theory, the causal model account predicts no cue selection in diagnostic situations where the opposite interpretation is made, while the associationist theory does. Note that we use the terminology "cause–effect" (CE) and "effect–cause" (EC) for "predictive" and "diagnostic" tasks, respectively.

The crucial question, then, is whether cue selection occurs in EC tasks. Waldmann and Holyoak reported an experiment (Experiment 3), which on the face of it seems to cast doubt on the associative theory's prediction that selection will occur under such circumstances. Subjects were told to imagine they were in a bank and had to learn relationships between the state of activation of the alarm system and various buttons (CE condition) or indicator lights (EC condition). Both tasks took place across two stages and used a blocking design. In Stage 1, a predictive cue P was consistently paired with the outcome in that it was present every time the outcome was present and absent whenever the outcome was absent. Another two cues were also presented, cue C, which occurred on every trial regardless of whether the outcome was present or not, and cue U, which was uncorrelated with the outcome. During Stage 2 the same cues were again present but a target cue R was added. This redundant cue was present only on those trials in which cue P was also present. Waldmann and Holyoak reasoned that since cues P and R had each been paired with the outcome, any

difference between them in terms of subjects' ratings at the end of Stage 2 would be evidence of cue selection. The relevant result was that an attenuation of the ratings of cue R occurred in the CE but not the EC context, so that, apparently, cue competition had arisen only in the former context, in accordance with the causal model theory and contradictory to an associationist analysis.

Waldmann and Holyoak's results are, on the face of it, highly problematic for associationist explanations of cue selection. Associationist accounts assume that when the subject is asked to predict an outcome, the cues represent the input and the outcome represents the output, with prediction responses being determined by unidirectional weights from the cues to the outcomes. Cue selection should emerge whenever multiple cues predict a single outcome, regardless of the possible causal order interpretation that can be put on the events. Since Waldmann and Holyoak's (1992, Experiment 3) result contradicts this prediction, our next study was undertaken in an attempt to examine their findings.

A. Experiment 2

We (Shanks & Lopez, 1996, Experiment 3) adopted a different experimental design to distinguish between the two accounts (see Table I). In this EC design, the trial types for the noncontingent cue A are AB $\rightarrow O_1$, B $\rightarrow O_1$, C $\rightarrow$ no O, and the trial types for the contingent cue D are DE $\rightarrow O_2$, E $\rightarrow$ no O, F $\rightarrow O_2$, where O_1 and O_2 refer to the outcomes and no O is no outcome. Applying the causal model contingency formula to the target cues A and D yields a value of .5 in each case:

$$\Delta p = p(\mathrm{E/C}) - p(\mathrm{E/-C}) = .5 - 0 = .5.$$

For instance, the AB $\rightarrow O_1$, B $\rightarrow O_1$, and C $\rightarrow$ no O trials are now interpreted as $E_A E_B \leftarrow \mathrm{cause}_1$, $E_B \leftarrow \mathrm{cause}_1$, and $E_C \leftarrow$ no cause trials, from which the crucial $\mathrm{cause}_1 \rightarrow E_A$ contingency is readily seen to be $\Delta p = .5$. The calculation is the same for the contingent cue D.

TABLE I

Trial Types in Experiment 2

	Cues → outcomes	Target relationship
Noncontingent	AB → O_1, B → O_1, C → no O[a]	A → O_1
Contingent	DE → O_2, E → no O, F → O_2	D → O_2

[a] A–F are symptoms; O_1 and O_2 are diseases; no O, no disease.

Thus, Waldmann and Holyoak's causal model theory predicts the absence of any difference between ratings for the contingent and noncontingent cues in this design. What about associationist theories? The addition of the extra trial type (C → no O and F → O_2) compared to Saavedra's design to equate the outcome frequencies does not effect the predictions of the Rescorla–Wagner (1972) theory, since the contingent cue remains a better predictor of the outcome than the noncontingent cue, and a selection effect is again predicted (see Shanks, 1991, p. 438, for a simulation of this trial design). The inclusion of the C → no O trials does not alter the fact that cue B is able to "overshadow" cue A and prevent it (at asymptote) from acquiring any associative strength for outcome O_1. Similarly, the inclusion of the F → O_2 trials does not impair cue D's ability to obtain an asymptotic associative strength of λ for outcome O_2. The experiment therefore provides a crucial test between the two theories. Note that Shanks (1991, Experiment 2) conducted an experiment with this design that yielded a significant difference in judgments for the target contingent and noncontingent cues. But although that study used an EC medical diagnosis task, the cues (symptoms) were given real names (e.g., swollen glands). Waldmann and Holyoak (1992) argued that subjects in this case may have recruited real-world knowledge of extraexperimental causal factors that would have altered the nature of the task. The present experiment therefore uses abstract cues such as "symptom A."

The cover story stated that on each trial a patient's symptoms would be described and the task was to predict the accompanying disease. The symptoms were labeled with letters (A to L) and diseases with numbers (1 to 4), and the specific letters were initially assigned at random to the trial types given in Table I. The cues appeared on the screen in a list, and for trial types containing more than one cue, the order of cues in the list was chosen at random. The subject selected a category response and typed the appropriate key, and then corrective feedback was given. After this training phase, subjects were given a questionnaire in which they had to rate on a scale from 0 to 100 how strongly cues A and D were associated with each of the outcomes.

The critical result is that contingent ratings (M = 66.6) were reliably higher than noncontingent ones (M = 58.2). Although the difference between these means is small, the results plainly show that a cue-selection effect occurs with this design. The results are at variance with the idea that subjects compute contingency in the manner prescribed by Waldmann and Holyoak's (1992) causal model theory. In contrast, the results are exactly what would be predicted by an associative account of learning such as the Rescorla–Wagner theory. One such a theory, the cues—regardless of their real-world interpretation—represent the inputs to an associative network