Journal of Experimental Psychology:
Learning, Memory, and Cognition
1993, Vol. 19, No. 6, 1398-1410

0278-7393/93/$3.00

OBSERVATIONS

Cue Competition in Human Categorization: Contingency or the Rescorla–Wagner Learning Rule? Comment on Shanks (1991)

Eric R. Melz, Patricia W. Cheng, Keith J. Holyoak, and Michael R. Waldmann

Shanks (1991) reported experiments that show selective-learning effects in a categorization task, and presented simulations of his data using a connectionist network model implementing the Rescorla–Wagner (R–W) theory of animal conditioning. He concluded that his results (a) support the application of the R–W theory to account for human categorization, and (b) contradict a particular variant of contingency-based theories of categorization. We examine these conclusions. We show that the asymptotic weights produced by the R–W model actually predict systematic deviations from the observed human learning data. Shanks claimed that his simulations provided good qualitative fits to the observed data when the weights in the networks were allowed to reach their asymptotic values. However, analytic derivations of the asymptotic weights reveal that the final weights obtained in Shanks' Simulations 1 and 2 do not correspond to the actual asymptotic weights, apparently because the networks were not in fact run to asymptote. We show that a contingency-based theory that incorporates the notion of focal sets can provide a more adequate explanation of cue competition than does the R–W model.

Shanks (1991) described three experiments in which subjects were asked to play the role of medical diagnosticians. After being presented with a series of case histories (patterns of patients' symptoms associated with various fictitious diseases), subjects were asked to rate how strongly they associated each symptom with each disease, using a 0–100 rating scale. Subjects' association ratings consistently varied with the relative predictiveness of each symptom for the disease as defined by the Rescorla and Wagner (1972) model (hereinafter the R–W model) rather than with the *cue validity* of the symptoms (i.e., the probability of the disease given a symptom; Shanks, 1991, Experiments 1–3) or with the *contingency* of the symptoms (i.e., the difference between the probability of the disease given the presence of a symptom and that probability given the absence of the symptom; Shanks, 1991, Experiments 2–3).

For example, consider the design of Shanks's (1991) Experiment 2. (Summaries of the experimental designs of Experiments 1–3 are presented in Table 1.) In Shanks's (1991) contingent set, the Compound Symptom AB signals the presence of Disease 1, but Symptom B alone signals the absence of the disease, whereas Symptom C alone signals the presence of the disease. In the noncontingent[1] set, Compound Symptom DE signals the presence of Disease 2, Symptom E alone also signals the presence of the disease, but Symptom F alone signals the absence of the disease. The critical comparison is between the association rating given to Symptom A for Disease 1 and the association rating given to Symptom D for Disease 2. The contingency computed over the entire set of events presented is .8 for both relations (see Figure 1); however, the R–W model predicts that D, which is paired with a better predictor, E, should be rated as less associated than the corresponding Symptom A, which is only paired with a nonpredictor, B. This difference was observed. In other words, the rating given to a cue was reduced if a competing cue was a better predictor of the relevant disease. Such cue competition, in which associative learning to one cue is blocked by the learning that accrues to a more predictive cue, has some similarity to results obtained in animal conditioning experiments.[2]

Eric R. Melz, Patricia W. Cheng, and Keith J. Holyoak, Department of Psychology, University of California, Los Angeles; Michael R. Waldmann, Department of Psychology, Universität Tübingen, Tübingen, Federal Republic of Germany.

Preparation of this article was supported by National Science Foundation Grant DBS 9121298 to Patricia W. Cheng and by a University of California, Los Angeles Academic Senate Research Support Grant to Keith J. Holyoak. Douglas Hintzman, Robert Nosofsky, and Thomas Wickens provided valuable comments on earlier drafts.

Correspondence concerning this article should be addressed to Patricia W. Cheng, Department of Psychology, University of California, Los Angeles, California 90024-1563. Electronic mail may be sent to cheng@cognet.ucla.edu.

[1] Because the critical cues in Shanks's (1991) noncontingent conditions were contingently related to the respective diseases by the conventional definition, the labels for his stimulus sets in Experiments 1 and 2—contingent condition and noncontingent condition—do not conform to conventional usage.

[2] In the animal conditioning literature the term *blocking* is reserved for a paradigm in which the animal is first conditioned to a single cue presented alone, which then blocks subsequent learning to a second cue that is always paired with the first cue when reinforcement is given. In this article we use the term blocking in a more general sense to refer to reduction in associative learning to

Table 1
Conditions, Trial Types, Number of Trials, and Percentage of Correct Diagnoses for Shanks's (1991) Experiments 1–3

Experiment/condition	Trial type	No. trials	% correct
Experiment 1			
Contingent	AB → D1	10	88
	B → 0	10	81
Noncontingent	CD → D2	10	75
	D → D2	10	88
Contingent	EF → D3	10	75
	F → 0	10	94
Noncontingent	GH → D4	10	88
	H → D4	10	69
Contingent	IJ → D5	10	75
	J → 0	10	81
Noncontingent	KL → D6	10	88
	L → D6	10	94
Experiment 2			
Contingent	C → D1	15	100
	AB → D1	15	100
	B → 0	15	94
Noncontingent	DE → D2	15	100
	E → D2	15	100
	F → 0	15	94
Experiment 3			
Correlated	AB → D1	20	91
	AC → 0	20	82
Uncorrelated	DE → D2	10	49
	DE → 0	10	49
	DF → D2	10	52
	DF → 0	10	52

Note. From "Categorization by a Connectionist Network" by D. R. Shanks, 1991, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* Tables 1, 2, and 3; pp. 436, 438, and 439, respectively. Copyright 1991 by the American Psychological Association, Inc. Adapted by permission.

Shanks (1991) interpreted his experimental results in terms of a connectionist categorization model that is based on an application of the R–W learning rule, which has been used to account for cue-competition effects in studies of animal conditioning. For all experiments, the results of the simulation model after training were qualitatively consistent with subjects' ratings when we compared the final weights of cue–disease associations in the network were compared with the human data. Table 2 shows subjects' associative ratings and the corresponding terminal weights obtained by Shanks's (1991) network. For example, in his Experiment 1, the mean associative rating was 62.3 for contingent cues and 41.8 for noncontingent cues. The comparable means for Experiment 2 were, respectively, 58.6 and 33.8. The terminal weights of the simulation for contingent and noncontingent cues were, respectively, 55.0 and 23.7 (Experiment 1) and 61.1 and 17.4 (Experiment 2). Thus, not only did the model correctly predict the ordinal relationship between ratings for the two types of cues, but it also apparently accounted for the fact that blocking of the less predictive cue was always partial rather than complete. Just as none of the mean association ratings were close to 0, no relevant terminal weight went to 0. This impressive fit between the data and the model's predictions therefore seems to provide strong support for an associationistic account of category learning. Shanks (1991) further concluded that his results cannot be explained by any of a number of alternative theories. In particular, he argued that cue validity and contingency theory as applied to the entire set of events cannot account for the results of these experiments.

An important qualitative aspect of Shanks's (1991) results in his Experiments 1 and 2 is that blocking of learning to redundant cues was only partial, rather than complete. Although subjects consistently gave higher association ratings to the more predictive cues, they also clearly gave the less predictive cues ratings indicative of a nonnegligible relationship to the disease. We show that, contrary to Shanks's (1991) claims, his connectionist model, in fact, does not predict the partial blocking observed in his Experiments 1 and 2. Moreover a contingency-based theory of causal induction and categorization may well account for cue competition. Furthermore, unlike Shanks's (1991) model, such a model potentially predicts the observed partial blocking.

Does the R–W Learning Rule Predict Partial Blocking at Asymptote?

In the past few years, there has been a surge of interest in the potential applicability of animal conditioning models to human categorization and causal induction. The R–W model and extensions of it, often implemented as connectionist networks, have figured prominently in these efforts (e.g., Gluck & Bower, 1988; Shanks, 1990). These extensions to higher level human learning have been advocated even though the R–W model has a number of well-known limitations as an account of animal conditioning. For example, the model is unable to account for learned irrelevance (the reduced conditionability of a cue as a result of an initial period of nonreinforcement of that cue) or for the conditions under which inhibitory cues can be extinguished (see Gallistel, 1990; Holland, Holyoak, Nisbett, & Thagard, 1986; Miller & Matzel, 1988). The present analyses reveal that the explanatory power of the R–W model as an account of human causal induction may be limited in additional ways.

Shanks (1991) indicated that the terminal strengths he reported are an inevitable consequence of the properties of the network:

> No systematic search of the parameter space was performed; however, the pattern of results is not due simply to the choice of a particular set of parameters. All the parameters really affect is the rate at which the associative strengths reach asymptote. The parameters chosen are such that the terminal associative strengths are at asymptote. (p. 436)

This statement strongly implies that the terminal associative strengths obtained by the network simulations are in fact the theoretical asymptotic weights of the network and that the simulation parameters have little or no effect on the terminal associative strengths.

a cue as a consequence of learning that accrues to a more predictive cue.

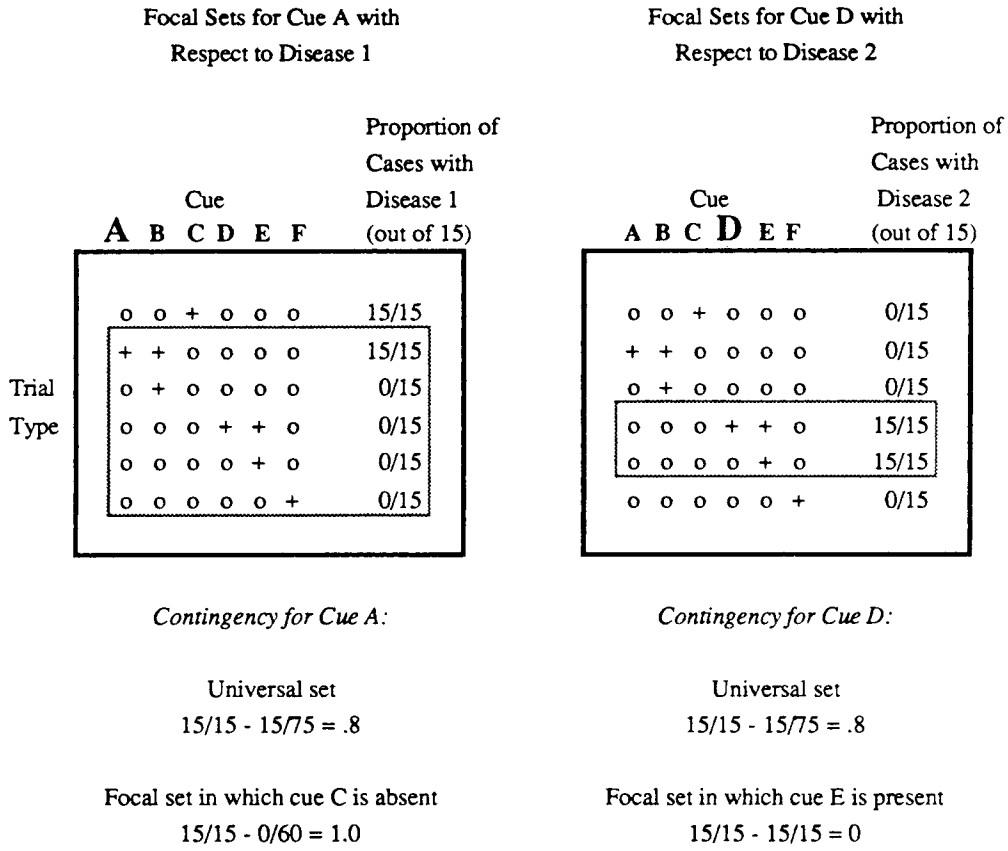

Figure 1. Potential focal sets in Shanks's (1991) Experiment 2. (Letters A to F denote cues. Solid-line rectangles indicate universal focal sets; dashed-line rectangles indicate conditional focal sets. Large bold letters denote the cues crucial for comparison.) From "Categorization by a Connectionist Network" by D. R. Shanks, 1991, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, Tables 1, 2, and 3; pp. 436, 438, and 439, respectively. Copyright 1991 by the American Psychological Association, Inc. Adapted by permission.)

In fact, given that Shanks's (1991) predictions are based solely on the asymptotic weights of the R–W model, it is not necessary to run computer simulations: The asymptotic weights can be calculated by analytic methods. In the Appendix, we derive the asymptotic weights for Shanks's (1991) simulations using a least-mean-squares method.[3] We found that, contrary to the simulation results reported by Shanks (1991), in his Experiments 1 and 2, the asymptotic weight for a contingent cue is the maximum value, 100, whereas the asymptotic weight for a noncontingent cue is 0. That is, the R–W learning rule predicts that at asymptote blocking in these experiments will be complete rather than partial. In the case of Experiment 3, our analysis indicates that the relative weights for different cues will vary with the choice of initial weights and with the learning parameters of the network.

Simulations of Experiments 1 and 2

Because the theoretical asymptotic weights differ radically from the terminal weights reported by Shanks (1991), we tried to understand the discrepancy by implementing network models that are based on Shanks's descriptions of his simulations and by comparing them with alternative simulations. Table 2 shows the terminal associative strengths for each simulation reported by Shanks (1991) along with our replication of the simulations conducted with the same parameter values and number of training trials as those he reported. For the simulations of Experiments 1 and 2, we closely replicated the terminal weights reported by Shanks (1991). However, on the basis of our theoretical analysis, it is clear that these terminal weights are not asymptotic. We therefore repeated all simulations, this time increasing the number of trials in each simulation by a factor of 20. As Table 2 indicates, these runs produced terminal weights extremely close to the theoretical asymptotic weights.

Simulation of Experiment 3

Following Rescorla and Wagner (1972), Shanks (1991) pointed out that the network correctly predicts the results in Experiment 3 only when the learning rate for reinforced trials

[3] There are alternative methods for deriving the asymptotic weights (e.g., Appendix A of Gluck & Bower, 1988).

Table 2
Cues, Associative Ratings, Terminal Network Weights From Shanks (1991), Replicated Terminal Weights, and Terminal Weights Obtained by Increasing the Total Trials by a Factor of 20

Experiment/cue	*M* rating	Shanks's reported weight	Replication weight		Extended replication	
			M	*SD*	*M*	*SD*
Exp. 1						
A	62.3	55.0	54.8	1.2	99.8	0.0
B	—	—	37.5	2.8	0.2	0.0
C	41.8	23.7	24.0	6.1	0.0	0.0
D	—	—	87.3	1.1	100.0	0.0
Exp. 2						
A	58.6	61.1	61.1	1.6	100.0	0.0
B	—	—	33.1	2.6	0.0	0.0
C	—	—	92.7	0.0	100.0	0.0
D	33.8	17.4	22.0	5.2	0.0	0.0
E	—	—	91.8	1.4	100.0	0.0
F	—	—	0.0	0.0	0.0	0.0
Exp. 3						
A	32.3	33.5	42.6	0.9	33.3	0.0
B	87.5	66.4	55.4	0.5	66.7	0.0
C	13.5	−32.9	−12.8	1.1	−33.3	0.0
D	49.0	59.3	60.0	2.8	60.5	2.8
E	39.2	29.3	30.1	2.2	30.1	2.0
F	43.8	30.1	30.0	1.9	30.4	2.0

Note. Exp. = Experiment. Dashes indicate data are not available. From "Categorization by a Connectionist Network" by D. R. Shanks, 1991, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* Tables 1, 2, and 3; pp. 436, 438, 439, respectively. Copyright 1991 by the American Psychological Association, Inc. Adapted by permission.

is greater than that for unreinforced trials. Our analysis (see Appendix) also shows that the pattern of weights reported by Shanks (1991) is dependent on the initial weights provided to the network. Our simulation confirms our analysis.[4]

Implications of Apparent Partial Blocking

Our analyses thus reveal that for the designs used in Shanks's (1991) Experiments 1 and 2, the asymptotic performance of the model—contrary to the original report—predicts that perfect predictors ought to completely inhibit associative learning to redundant predictors. Furthermore, this behavior ought to be independent of parameter values or initial conditions, as our analysis in the Appendix shows. Neither Shanks's (1991) experiments nor similar studies (Chapman, 1991; Chapman & Robbins, 1990; Dickinson, Shanks, & Evenden, 1984; Shanks, 1985; Waldmann & Holyoak, 1992) have yielded complete blocking in human subjects.

It might be argued that a positive rating, even one that is significantly greater than 0, does not necessarily indicate a positive association for Experiment 1, because in that experiment subjects apparently were not given a clear definition for the endpoints of the scale. Thus, subjects might not have interpreted a rating of 0 as indicating no association. In Shanks's (1991) Experiment 2, however, the zero point was defined. Experiment 2 subjects were told, "Zero indicates that the symptom and disease are completely unrelated, and 100 indicates that they are very strongly related" (Shanks, 1991, p. 438).

Subjects' interpretations of the zero point can be tested by comparing the association rating between a disease and a redundant symptom, such as that between Disease 2 and Symptom D in Experiment 2, with that between a disease and a completely irrelevant symptom, such as that between Disease 1 and Symptom D in the same experiment. The R–W rule predicts that both kinds of associations should have zero weight. If the association ratings observed for redundant symptoms proved greater than those for irrelevant symptoms, then subjects' interpretations of the rating scale would be ruled out as an explanation of the apparent partial blocking of redundant Cue D. Unfortunately, although Shanks (1991) included both redundant and irrelevant symptoms in his designs for Experiments 1 and 2, he did not report the relevant ratings needed for a comparison.

The equality of these two kinds of associations strikes us as rather implausible. In a test of a similar prediction of the R–W rule, Waldmann and Holyoak (1992, Experiment 3) asked subjects to rate the degree to which each of several individual cues was predictive of an effect, using a scale from 0 to 10, where 0 indicated the cue was not a predictor and 10 indicated that the cue was a perfect predictor. The cues that were rated included one that was a perfect predictor of the effect, one that was redundant with the predictor (i.e., subject to blocking), one that was constantly present, and one that varied in a noncontingent manner. According to the R–W

[4] Our replication of Shanks's (1991) simulation of Experiment 3 following his description failed to reproduce his results. Specifically, the terminal weights of Cues A, B, and C are significantly different from those that Shanks reported. However, by increasing the number of training trials by a factor of 20 beyond that reported by Shanks, we obtained terminal weights similar to those he reported (see Table 2).

model, the associative strengths acquired for the latter three cues all should have approached 0. However, Waldmann and Holyoak found that the mean of subjects' final predictiveness ratings was significantly higher for the redundant cue relative to the two uncorrelated cues: mean ratings of 4.3 for the redundant cue, 1.3 for the constant cue, and 0.8 for the varying noncontingent cue. The partial blocking obtained by Waldmann and Holyoak, like that observed in Shanks's (1991) experiments, raises the possibility that humans are sensitive to differences in the causal status of cues that cannot be distinguished by the R–W model.

An alternative explanation of the partial blocking observed in Shanks's (1991) experiments is that the subjects were not in fact trained to asymptote. The subjects in Shanks's (1991) Experiment 2 were close to 100% correct for each trial type; however, their reaction time might well continue to shorten with increased training, indicating further learning. The evidence for partial blocking is thus not definitive. However, given that the R–W model makes a strong prediction of complete blocking at asymptotic learning, whereas only partial blocking has been reported for human causal induction, it seems that the burden of proof rests on proponents of the model to provide a convincing case of complete blocking at asymptote.

Contingency-Based Theories of Categorization and Causal Induction

Associative learning models are often contrasted with models that are based on statistical contingency. If there is only one possible causal factor, its contingency with an effect is defined as the difference between the probability of the effect given the presence of the factor and that probability given its absence. When multiple potential causal factors are present, there are various possible ways of computing contingency. Shanks (1991) and others (e.g., Chapman & Robbins, 1990; but see Chapman, 1991) have examined the special case in which the contingency of each of the multiple potential causal factors that are present is calculated independently (e.g., unconditionally over the universal set of events). However, contingency could alternatively be computed over subsets of the universal set of events. Philosophers and statisticians have proposed that in such situations causal relations should not be based on contingencies computed over the universal set of events (Cartwright, 1979, 1989; Reichenbach, 1956; Salmon, 1980, 1984; Simpson, 1951; Suppes, 1970, 1984) because in these situations unconditional contingencies do not reflect what people intuitively judge to be normative causal inferences. In particular, people distinguish between a genuine cause and a spurious cause—a factor that is contingently related to the effect purely as the result of its correlation with a genuine cause. Unconditional contingencies do not reflect this difference.

The proposal, then, is that if a factor is known or likely to be the cause of an effect, determining the causal status of another factor requires that the contingency of the latter be calculated separately conditional on the presence and on the absence of that cause (a test of "conditional independence").[5] Testing for conditional independence is analogous to using control conditions in standard experimental design, in which extraneous variables are kept constant. Although this criterion has not been uncontested among philosophers (e.g., Cartwright, 1989; Salmon, 1984), the prevalent adoption of the analogous principle of experimental design gives an indication of its normative appeal.

Let us consider the interpretation of some possible outcomes of the test of conditional independence for a target factor that is statistically relevant to the effect in terms of its unconditional contingency. For example, suppose we are assessing possible causes of cancer and that smoking cigarettes is an established cause. Now we observe that coffee drinking is also statistically relevant to cancer, in that the probability of cancer is higher for people who drink more than five cups per day than for those who drink fewer cups per day. However, let us further suppose that people who drink large quantities of coffee also tend to smoke. To tease apart the influence of coffee drinking from that of smoking, it is desirable that we calculate the conditional contingency between coffee drinking and cancer separately for cases involving the presence versus the absence of smoking. The following are four possible outcomes that will be relevant in interpreting Shanks's (1991) results:

Case 1. If both conditional contingencies for the target factor are positive, then the target factor will be interpreted as a genuine cause. For example, if coffee drinking increases the risk of cancer both for smokers and for nonsmokers, then coffee drinking will be interpreted as a genuine cause (unless it turned out to be confounded with some other cause of cancer, such as eating fatty foods).

Case 2. If contingencies for the target factor conditional on both the presence and the absence of the established cause are zero, then that factor will be interpreted as a spurious cause. It is said to be "screened off" (i.e., normatively blocked) from the effect by the conditionalizing cause. For our example, the statistical link between coffee drinking and cancer would be attributed entirely to the confounding between coffee drinking and smoking.

Case 3. If the effect always occurs in the presence of the established cause, regardless of whether the target factor occurs (therefore the contingency conditional on the presence of the established cause is zero), but the contingency conditional on the absence of the causal factor is positive, then the target factor will be interpreted as a genuine cause. This situation would arise if smoking always caused cancer, so that coffee drinking did not increase the risk of cancer for smokers but did increase the risk for nonsmokers. In this situation, coffee drinking would be interpreted as a genuine cause of cancer, and the zero contingency in the presence of smoking would likely be attributed to a ceiling effect (i.e., smoking by itself generates the maximal cancer risk, so that the detrimental impact of coffee drinking is masked for smokers).

Case 4. If the contingency of the target factor conditional on the presence of the established cause is positive, but the effect never occurs in the absence of the established cause (therefore the contingency conditional on the absence of the known cause is zero), then the two factors will be interpreted as interacting to produce the effect (see Cheng & Novick,

[5] We do not mean to imply that at least one of the two conditional contingencies can always be computed, nor that a test of conditional independence is the only process for differentiating between genuine and spurious causes (Lien & Cheng, 1992).

1992). Such an interaction would exist if coffee drinking interacted with smoking to increase the risk of cancer for smokers but had no effect on the probability of cancer for nonsmokers.

What are the implications of the various ways of computing contingency for Shanks's (1991) results? One problem that complicates the test of conditional independence is that the information required for computing the two conditional contingencies is not always available. In the classical blocking paradigm, for example, one cue is established as a perfect predictor of the unconditioned stimulus (US) in the first phase of the learning task (Kamin, 1969). In a second phase, this cue and a redundant second cue are, in combination, always paired with the US. Across the two phases, the contingency of the redundant cue conditional on the presence of the predictive cue is therefore zero. Because the redundant cue is never presented by itself, it is impossible to compute the contingency between this cue and the US, conditional on the absence of the predictive cue. As Waldmann and Holyoak (1992) noted, because the level of the effect produced by the predictive cue is already at ceiling, it is impossible to determine whether the redundant cue is a spurious cause (Case 2), or a genuine cause (Case 3). Given that relevant information is missing in the blocking design, subjects who adopt the criterion of conditional independence would be uncertain about the predictive status of the redundant cue, as opposed to being certain that this cue is not predictive, as implied by the R–W learning rule.

In some other situations the contingency for a potential causal factor conditional on the presence of a likely or known alternative cause is not available. In such situations, when the contingency for this factor, conditional on the absence of the cause, is zero, one cannot be certain whether the factor is a spurious cause (Case 2) or interacts with the alternative factor as a conjunctive cause (Case 4).

It is important to note that there is an asymmetry between the informativeness of tests conditional on the absence versus presence of other causes: The tests most likely to clearly rule out a target factor as an independent excitatory cause are those that are based on the absence of conditionalizing cues. In particular, the finding of a zero contingency conditional on the absence of other causes clearly rules out the factor as an independent excitatory cause (i.e., it is either spurious as in Case 2, or a component of an interaction as in Case 4), whereas the finding of a zero contingency conditional on the presence of a known cause is inconclusive (the target might be spurious as in Case 2, but it might instead be genuine as in Case 3). (This interpretation excludes consideration of inhibitory causes.) Similarly, finding a positive contingency conditional on the absence of other causes constitutes evidence that the cue is an independent excitatory cause (for which Case 1 or Case 3 might obtain), but a positive contingency conditional on the presence of a known cause could indicate either a genuine independent excitatory cause (as in Case 1) or a component of an interactive excitatory cause (as in Case 4). The fact that tests conditional on the absence rather than the presence of other causes are more informative is reflected in experimental design: If only one type of conditionalizing test can be performed, scientists generally favor designs in which a target factor is manipulated while ensuring that other known causes are absent, rather than present. We therefore assume that people will prefer to conditionalize each target factor on the simultaneous absence of all established or likely causes, because this is the test that will be maximally informative.

Contingency models focus on reasoning from causes to effects. However, the approach can also be applied to situations in which cues are interpreted not as causes per se, but as causal indicators that predict an outcome. In Shanks's (1991) experiments, for example, it seems implausible to suppose that his subjects actually interpreted the symptoms as causes of the diseases; however, they might have treated the symptoms as causal indicators (i.e., visible correlates of some hidden cause of a disease, such as a virus). From Shanks's (1991) report, it appears that in all three experiments his subjects were told, "All you have to do is to learn which symptoms tend to go with which illness, so that you can make as many correct diagnoses as possible" (p. 436). During a learning trial, subjects were first told the symptoms of a patient, and then they were asked to judge from which disease the person was suffering. Feedback followed the subject's judgment. In Experiments 1 and 2, the rating task instructions said, "On a scale from 0 to 100, how strongly do you associate symptoms [S] with disease [D]?" (Shanks, 1991, p. 436). Given the vagueness of the instructions, at least some subjects presumably treated the symptoms as causal indicators of the diseases and encoded probabilities of each disease conditional on the symptoms. Although the symptoms may have been interpreted as causal indicators rather than actual causes, for simplicity we will refer to the symptoms as "causes" and the diseases as "effects." In our discussion of the contingency analysis of Experiment 2, we will consider the implications of the further possibility that some subjects may have interpreted the symptoms as effects of the diseases.

A Process Model for Assessing Conditional Independence

To provide a more detailed analysis of Shanks's (1991) results in terms of contingency theory, we apply a contingency-based process model developed by Cheng and Holyoak (in press). This proposed process model is of course only one possible procedure for contingency analysis, and we do not mean to deny the viability of other variants. This model, however, will serve to illustrate how a contingency analysis might account for results in Shanks's (1991) experiments.

A plausible psychological model of causal inference that is based on contingency analysis must specify mechanisms that would allow people to decide (a) what cues should be used to conditionalize others, (b) what conditional tests to perform once a set of conditionalizing cues has been selected, and (c) how to integrate the resulting contingency information to make causal assessments of the cues. In situations in which there is no guidance from prior knowledge (as in Shanks's, 1991, experiments), every cue is potentially causal. Given n binary cues, exhaustively conditionalizing the contingencies for each target cue on every combination of the presence and absence of the other cues require computing

$2^{n-1} \cdot n$ contingencies. Given processing limitations, it is crucial to specify how people select which contingencies to compute. Moreover, in Shanks's (1991) experiments many of the cue combinations that would be relevant to a contingency analysis were never presented to subjects. Accordingly, it is necessary to specify which contingencies will be computed in the face of missing information.

Let us first consider the selection of conditionalizing cues. The ideal set of conditionalizing cues would include all, and only, those that are actually causal. Given the limitations of knowledge, the best people could do is to select as conditionalizing cues those that they currently believe to be plausible causes. In cases in which prior knowledge is relevant, such knowledge would be used to establish certain cues as likely causes, and the contingencies for other cues would then be conditionalized on the (perhaps tentatively) established causes. If such prior knowledge is lacking, people may nonetheless use some heuristic criterion to select an initial set of conditionalizing cues. A simple heuristic that might be used is to include any cue that is noticeably associated with the effect, based on the relative frequency of the effect given the cue. Contingencies are not computed in this initial phase of selecting conditionalizing cues; rather, people simply identify a pool of cues with some apparent positive association with the effect, which are then treated as an initial set of plausible causes. This phase ignores the possibility that cues may be interactive or inhibitory causes.

Contingency assessment occurs in the subsequent phase, in which people compute the conditional contingencies of all cues on the basis of the set of conditionalizing cues identified in the initial phase. The initial set of conditionalizing cues can be dynamically updated if contingency assessments indicate that cues that at first appeared to be plausible causes are in fact spurious or that cues initially viewed as causally irrelevant are in fact causal. Changes in the set of conditionalizing cues in turn change the relevant conditional contingencies for all cues, which may alter other causal assessments. The entire assessment process thus may be iterative. If the values of the cues stabilize as the process iterates, the process will return these values and stop. Otherwise, the process will stop after some maximum number of iterations.

In assessing conditional contingencies, heuristics are required to determine which tests (of those possible, given the cue combinations that are actually presented) should in fact be performed. We assume, on the basis of the arguments presented earlier, that people prefer to conditionalize the contingency for each target factor on the simultaneous absence of all conditionalizing cues. If this is not possible, then they will conditionalize on the absence of as many conditionalizing cues as possible. For any conditionalizing cue that cannot be kept absent along with the others, it is desirable to conditionalize the contingency for the target factor on its presence (holding other conditionalizing cues absent).

A special case that should be noted involves any cue that is constantly present (i.e., part of the background context). In applications of the R–W model to conditioning phenomena, it is commonly assumed that constant background cues are represented and may acquire nonzero weights. Note that it is impossible to compute any contingencies at all (either conditional or unconditional) for constant cues because they are always present. Accordingly, subjects will have no positive evidence that any constant cue is causal. Therefore although constant cues may be initially included in the set of conditionalizing cues (because of their presence when the effect occurs), the failure to obtain contingency information to support their causal status will lead to them being dropped. Because consideration of constant cues would not alter our contingency analyses for Shanks's (1991) experiments, we will not consider them further.

In addition to specifying what cues are selected to form the conditionalizing set and what contingencies are computed, a process model must specify a response mechanism that translates the calculated contingencies into causal judgments. If the experimental design omits cases that would be relevant in assessing conditional independence for a target factor, subjects may find themselves uncertain about its causal status after considering conditional contingencies. At least in such cases, subjects may base their causal assessments on the unconditional as well as (or instead of) on the conditional contingencies for cues. Mean ratings across subjects may therefore reflect some mixture of the evidence provided by conditional and unconditional contingencies.

Following Cheng and Novick's (1990) terminology, we call the set of events over which a subject computes a particular contingency[6] a *focal set*. When subjects do not all use one and the same focal set, the mean causal judgment about a cue (averaged across subjects in an experimental condition) should reflect some mixture of assessments that is based on the multiple focal sets used. These may include the universal focal set[7] of all events in the experiment (i.e., unconditional contingencies) and various more restricted focal sets (i.e., conditional contingencies). The response mechanism must then account for the ways in which multiple contingencies are integrated. The clearest situation is that in which the relevant unconditional and conditional contingencies for a factor are all computable and equal to zero, in which case subjects should be certain that the factor is noncausal. Such cues should therefore receive mean ratings equal to or approaching zero in Shanks's (1991) Experiments 2 and 3, in which the zero point on the rating scale was unambiguously defined as indicating that the symptom and disease were completely unrelated. Beyond this limiting case, we make no claim about the exact quantitative mapping between contingency values and the numerical response scales used by subjects in Shanks's (1991) experiments. Our assumption is that subjects' causal estimates will increase monotonically with a nonnegatively weighted function of the contingency values for their focal sets. Individual subjects may calculate and integrate multiple contingencies for a cue (e.g., by simple averaging). Alternatively, each subject may use only one fo-

[6] Cheng and Novick (1990) referred to a contingency value as a "contrast" (between the conditional probability of the effect in the presence versus absence of the causal factor).

[7] What is referred to here as the "universal set" is actually the pragmatically restricted set of events that occur in the experiment (i.e., a small subset of the truly universal set of events known to the subject). This contextual delimitation of the largest relevant focal set implies that even the cases in the *cause-and-effect-both-absent* cell are restricted to a small finite number.

cal set, but different subjects may use different focal sets in which case the mean ratings may mask distributions that are in fact multimodal. We refer to the assumption that causal ratings may be based on multiple contingencies (calculated either by individual subjects or by different subjects) as the "mixture-of-focal-sets" hypothesis.

In summary, the process model described by Cheng and Holyoak (in press) assumes that subjects will (a) identify as initial conditionalizing cues those that are noticeably associated with the effect; (b) compute contingencies for each target factor, conditional on the absence of as many conditionalizing cues as possible, thus dynamically revising the set of conditionalizing cues in the process; and then (c) use the computed conditional contingencies and/or unconditional contingencies to produce causal assessments for the cues.

Contingency Analysis of Experiment 2

For Experiment 1, the unconditional contingencies of the Critical Cue A (1.0) in the contingent set and Cue C (.91) in the noncontingent set yield the prediction of lower associative ratings for C than A. As Shanks (1991) noted, this difference in contingency can explain cue competition. In addition, the positive contingency for C can predict the incomplete blocking of C. A contingency analysis using conditional contingencies, however, can also provide an explanation for the results of Experiment 2, in which the unconditional contingencies were equated across sets. We consider how this analysis applies to the design of this experiment (see Table 1), for which the R–W rule predicts the complete blocking of the redundant cue.

In the contingent set, the Compound Cue AB signals the presence of Disease 1 (15 trials), but Symptom C by itself does so as well (15 trials). However, Cue B by itself signals the absence of the disease (15 trials), as does the absence of A, B, and C (45 trials). In the noncontingent set, Compound Cue DE signals the presence of Disease 2 (15 trials), as does the presence of Cue E alone (15 trials). In contrast, Cue F alone signals the absence of the disease (15 trials), as does the absence of D, E, and F (45 trials).

Figure 1 illustrates the computation of contingencies for the crucial cues for comparison: A in the contingent set and D in the noncontingent set. As shown in Figure 1, the unconditional contingency (i.e., the contingency computed over the universal set of all events) is .8 for Critical Cue A with respect to Disease 1, as is that for Cue D with respect to Disease 2. To test conditional independence of these cues with respect to the particular disease, we apply the process model described earlier. With respect to Disease 1 (see lefthand diagram of Figure 1), only Cues A, B, and C will be identified as initial conditionalizing cues, because these are the only cues that are ever accompanied by Disease 1. Cue B has a contingency of 0 in the focal set in which both A and C are absent (Figure 1, Rows 3–6 of the lefthand diagram). Cue C has a conditional contingency of 1.0 in the focal set in which Cues A and B are both absent (Figure 1, Rows 1 and 4–6 of the lefthand diagram). Each of the remaining Cues (D, E, and F) has a conditional contingency of 0 in the focal set in which all conditionalizing cues (A, B, and C) are absent (Figure 1, Rows 4–6 of the lefthand diagram). The contingency for Cue A, conditional on the absence of both B and C, cannot be computed because A does not occur in the absence of B. However, A has a contingency of 1.0 in the focal set in which B is present and C is absent (Figure 1, Rows 2–3 of the lefthand diagram). From the first iteration of conditional-contingency assessment, it follows that B will be assessed as noncausal and dropped from the set of conditionalizing cues, so that only A and C remain as conditionalizing cues. The relevant contingency for A then becomes that which is conditional on the absence of C (Figure 1, Rows 2–6 of the lefthand diagram, enclosed by a dashed rectangle) and has a value of 1.0. This is equal to the value of the relevant conditional contingency obtained for A in the previous iteration. As is the case for A, none of the values of the relevant conditional contingencies for any of the other cues change as a result of dropping B from the conditionalizing set.

For Disease 2 (see the righthand diagram of Figure 1), Cues D and E will be selected as conditionalizing cues. Because D never occurs in the absence of E, its contingency can only be calculated as conditional on the presence of E. For this focal set (enclosed by the dashed rectangle in Figure 1), the conditional contingency for D with respect to Disease 2 is 0. The difference between the computed contingency for Cue A with respect to Disease 1 (1.0) and that for Cue D with respect to Disease 2 (0) provides an explanation for cue competition—lower ratings are given to D than to A.

In addition, Cue E has a contingency of 1.0 conditional on the absence of D (Figure 1, Rows 1–3 and 5–6 of the righthand diagram). All other cues have a contingency of 0 with respect to Disease 2 in the absence of Cues D and E.

Next, we consider how partial blocking might arise. The conditional contingency for Cue D seems to predict that D will be completely screened off (and hence blocked) by Cue E. However, subjects may be uncertain of this conclusion, as it was not possible to conditionalize the contingency for D on the absence of E. Accordingly, at least some subjects may assess the unconditional contingency for D over the universal set of events, which is .8. This positive value contrasts with the 0 contingency for D conditional on the presence of E. Assuming that subjects' causal ratings reflect a mixture (either within individual subjects or across subjects) of these two contingencies, D will receive a relatively low but positive mean rating. That is, Cue D will be partially blocked. Moreover, the prediction of cue competition remains, because the mixture of contingencies for A (.8 and 1.0) is still higher than the mixture of the contingencies for D (.8 and 0). In summary, if there is a mixture of focal sets, either within subjects or across subjects, contingency theory predicts partial blocking in addition to cue competition.

The analyses described above assume that subjects used the symptoms as predictors of the diseases. It is also possible, however, that some of Shanks's (1991) subjects interpreted the disease labels to be denoting causes of symptoms, in which case they may have interpreted the task as one involving diagnostic learning (Waldmann & Holyoak, 1992). In diagnostic inference, the symptoms would be viewed as effects, and the diseases would be interpreted as causes of the symptoms. There is evidence that when people impose a causal schema on a situation, they tend to encode knowledge

about conditional probabilities in the cause-to-effect direction (Tversky & Kahneman, 1980). This directional preference emerges even when the causal direction is opposite to the temporal presentation order of cues and responses (Waldmann & Holyoak, 1992). Waldmann and Holyoak have shown that the degree of cue competition differs radically depending on whether people interpret the cues as the causes of an effect to be predicted, or as the effects of a cause to be diagnosed. In their study, redundant cues competed against one another when they were interpreted as possible causes but not when they were interpreted as possible effects. If the same phenomenon occurred in Shanks's (1991) experiments, one would expect that subjects who interpreted the cues as effects would show no cue competition, that is, no difference between the critical symptoms in the contingent and noncontingent sets.

As we explained for Shanks's (1991) Experiment 2 (and explain below for Experiment 3), the same prediction (no cue competition) happens to follow for subjects who reasoned from symptoms to diseases with contingency computed over the universal set. Thus for Shanks's (1991) data, adding to the mixture-of-focal-sets hypothesis the assumption that some subjects interpreted the cues as effects yields predictions that are qualitatively indistinguishable from those of the mixture-of-focal-sets hypothesis alone. To tease apart these different theoretical possibilities, researchers in this area need to control the focal sets and assure that the subjects understand the directionality of causal connections among cues.

As we noted earlier, the results of Shanks's (1991) Experiment 1 can be explained in terms of unconditional contingencies alone. If subjects in that experiment applied the process model to compute conditional contingencies, the pattern of causal assessments likewise yields cue competition and partial blocking. The contingency analysis is highly similar to that for Experiment 2.

Contingency Analysis of Experiment 3

Next, we consider a contingency analysis of the design of Experiment 3 (summarized in Table 1). In the correlated set, the Compound Cue AB signals the presence of Disease 1 (20 trials). However, the Compound Cue AC signals the absence of Disease 1 (20 trials), as does the absence of A, B, and C (40 trials). In the uncorrelated set, Compound Cue DE signals the presence of Disease 2 half of the time, on 10 of 20 trials, as does the Compound Cue DF (20 trials). In contrast, the absence of D, E, and F signals the complete absence of Disease 2 (40 trials). Shanks (1991) reported that subjects rated the Critical Cue A lower than the Critical Cue D (see Table 2). In addition, Cue B was rated much higher than C.

Figure 2 illustrates some of the focal sets involved in a contingency analysis of this design. The unconditional contingency is .5 for Critical Cue A with respect to Disease 1 (see lefthand diagram of Figure 2), as is that for Critical Cue D with respect to Disease 2 (righthand diagram of Figure 2). With respect to Disease 1, the process model identifies Cues A and B as initial conditionalizing cues. In the absence of these conditionalizing cues, Cues D, E, and F have 0 contingencies with respect to this disease. For Cues A, B, and C, because of the inherent confoundings between them in the experimental design, the most informative contingencies that can be computed are those conditional on the absence or presence of one conditionalizing cue. Cue A and Cue C each have a conditional contingency of 0 in the absence of B (the focal set enclosed within the rectangle in the lefthand diagram of Figure 2). Cue B has a conditional contingency of 1.0 in the presence of A. Because the relevant conditional contingency for Cue A is 0, A will no longer be viewed as a plausible cause, and hence will be dropped from the set of conditionalizing cues. Cue B, the sole remaining conditionalizing cue, has an unconditional contingency of 1.0,

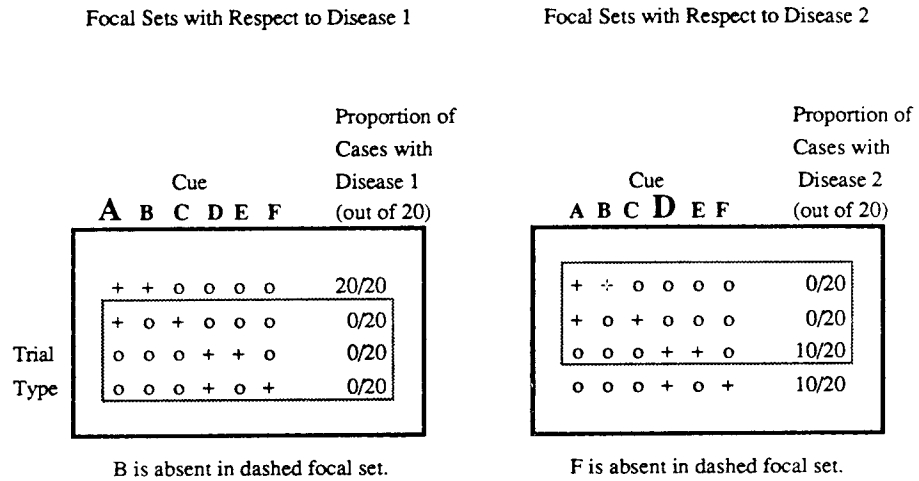

Figure 2. Potential focal sets in Shanks's (1991) Experiment 3. (Letters A to F denotes cues. Solid-line rectangles indicate universal focal sets, dashed-line rectangles indicate conditional focal sets. Large bold letters denote cues crucial for comparison. From "Categorization by a Connectionist Network" by D. R. Shanks, 1991, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, Tables 1, 2, 3; pp. 436, 438, and 439, respectively. Copyright 1991 by the American Psychological Association, Inc. Adapted by permission.)

Table 3
Unconditional and Conditional Contingencies, Dominance Contingency Ranking, Mean Contingency Ranking, and Mean Associativeness Ratings for Cues in Shanks's (1991) Experiment 3

Cue	Unconditional contingency	Conditional contingency	Dominance contingency ranking	Mean contingency ranking	Mean associativeness rating
A	0.5	0	3	4	32.3
B	1.0	1.0	1	1	87.5
C	–0.33	0	4	5	13.5
D	0.5	0.5	2	2	49.0
E	0.33	0.5	3	3	39.2
F	0.33	0.5	3	3	43.8

Note. The contingencies and ratings for Cues A, B, and C correspond to Disease 1; those for Cues D, E, and F correspond to Disease 2. From "Categorization by a Connectionist Network" by D. R. Shanks, 1991, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* p. 439. Copyright 1991 by the American Psychological Association, Inc. Adapted by permission.

confirming its status as a plausible cause. Thus the contingency analysis for Disease 1 provides evidence that only Cue B is a genuine cause of the disease. However, as much of the information for computing conditional contingencies is missing in this design, these conclusions might be viewed with uncertainty.[8] Recall that the unconditional contingency of A is .5. The mixture-of-focal-sets hypothesis therefore predicts that the association rating for A will be low but positive.

We now apply the same analysis to Disease 2 (see righthand diagram of Figure 2). Cues D, E, and F all qualify as conditionalizing cues. In the absence of these conditionalizing cues, A, B, and C each have a conditional contingency of 0. For Cues D, E, and F, given the inherent confoundings among these cues in the experimental design, the most informative contingencies that can be computed are those conditional on the absence of one conditionalizing cue. For example, the focal set enclosed by the rectangle in the right half of Figure 2 would be used to compute the contingency for the crucial D cue (as well as E), conditional on the absence of F. Similarly, the focal set conditional on the absence of Cue E (Figure 2, Rows 1, 2, and 4 of the righthand diagram) would be used to compute the contingency for Cues D and F. Every such conditional contingency, for Cues D, E, and F, has the value .5. By this analysis, these three cues are therefore equally likely to be viewed as causes of Disease 2. As was the case for Disease 1, because much of the information for computing conditional contingencies is missing, these conclusions might be viewed with uncertainty. For example, the conditional contingency for Cue E, in the focal set in which D is present, has no clear interpretation, because Cue F—which has been initially assessed as a plausible cause and is negatively correlated with Cue E in this set—cannot be disregarded. The interpretation of the conditional contingency for Cue F in the presence of D involves the same problem. As the unconditional contingency for Cue D is .5, the mixture hypothesis predicts that the association ratings will be higher for D (based on a mixture of contingencies of .5 and .5) than for A (based on a mixture of the contingencies 0 and .5), as reported by Shanks (1991).

Table 3 presents a summary of the contingency assessments for all the cases in Experiment 3 for which Shanks (1991) reported judgment data: Cues A, B, and C with respect to Disease 1 and Cues D, E, and F with respect to Disease 2. In Table 3 we report both the unconditional contingency for each case and the mean of the most relevant conditional contingencies that can be calculated. Table 3 also shows the predicted rank order of causal judgments derived by two methods that are based on the mixture hypothesis. The dominance method requires no assumptions about the relative weighting of the two contingencies for each case: A case is ranked higher than others only if at least one of its two contingencies is higher than the corresponding contingency for every dominated case and if neither of its contingencies is lower than the corresponding contingency for any dominated case. We obtained the second set of rankings by averaging the two contingencies for each case and assuming that they are weighted equally. It is apparent that both methods produce rank orderings that agree with the relative mean ratings that subjects gave for the six cases. (The comparable predictions derived from the R–W model appear in Table 2.)

Conclusion

In general, we find reasons to doubt whether variants of associationist models, which have encountered difficulty in dealing with various phenomena observed in studies of conditioning with animals, will be able to provide a full account of human categorization and causal induction. In any case, researchers of induction cannot afford to ignore the concepts of conditional independence and focal sets. Contingency theory may account for cue competition in human categorization tasks, once the role of conditional independence is recognized. Moreover, a contingency theory that specifies the role of focal sets offers a possible explanation for the partial blocking apparently observed in Shanks's (1991) Experiments 1 and 2 and in similar studies, findings which contradict the asymptotic predictions of the R–W model. Although current findings concerning partial blocking do not clearly discriminate between the two approaches, a number

[8] For example, a more complex, but nonetheless coherent, alternative interpretation that could be offered is that A is an excitatory cause, C is an inhibitory cause, and B might be either an excitatory cause or noncausal.

of other phenomena provide stronger evidence favoring contingency theory. In particular, contingency theory offers simple explanations of learned irrelevance, extinction of conditioned inhibition, and other phenomena that elude the R–W model (see Cheng & Holyoak, in press).

Our analysis of Shanks's (1991) results is of course post hoc and speculative, given that his experiments neither measured nor manipulated subjects' focal sets. Focal sets can, however, be measured and manipulated (see Cheng & Novick, 1990, 1991; Novick, Fratianne, & Cheng, 1992). Our purpose in this article was not to decide among alternative approaches but rather to introduce a variant of contingency theory that might challenge the R–W rule, to clarify the predictions of the R–W rule, and thereby to bring attention to previously ignored tests that may discriminate among these different approaches.

References

Cartwright, N. (1979). *How the laws of physics lie.* Oxford, England: Clarendon Press.

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, England: Clarendon Press.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 837–854.

Chapman, G. B., & Robbins, S. I. (1990). Cue interaction in human contingency judgment. *Memory & Cognition, 18,* 537–545.

Cheng, P. W., & Holyoak, K. J. (in press). Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In J.-A. Meyer & H. Roitblat (Eds.), *Comparative approaches to cognition.* 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.

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

Gallistel, C. R. (1990). *The organization of learning.* Cambridge, MA: MIT Press.

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

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 276–296). New York: Appleton-Century-Crofts.

Lien, Y., & Cheng, P. W. (1992). *How do people judge whether a regularity is causal?* Paper presented at the 33rd annual meeting of the Psychonomic Society, St. Louis, MO. November.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press.

Novick, L. R., Fratianne, A., & Cheng, P. W. (1992). Knowledge-based assumptions in causal attribution. *Social Cognition, 10,* 299–332.

Reichenbach, H. (1956). *The direction of time.* Berkeley, CA: University of California Press.

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.

Salmon, W. C. (1980). Probabilistic causality. *Pacific Philosophical Quarterly, 61,* 50–74.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

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

Shanks, D. R. (1990). Connectionism and the learning of probabilistic concepts. *Quarterly Journal of Experimental Psychology, 42A,* 209–237.

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

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B (Methodological), 13,* 238–241.

Suppes, P. (1970). *A probabilistic theory of causality.* Amsterdam: North-Holland.

Suppes, P. (1984). *Probabilistic metaphysics.* Oxford, England: Basil Blackwell.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88,* 135–170.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 49–72). Hillsdale, NJ: Erlbaum.

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.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Part 4,* 96–104.

Appendix

Derivation of Asymptotic Weights for Networks Used in Simulations by Shanks (1991)

The Rescorla–Wagner theory is described by Shanks (1991) as follows:

> On each trial, the weight connecting each feature that is present with the category occurring on that trial is incremented according to the equation
>
> $$\Delta V_{fc} = \alpha_f \cdot \beta_c \cdot (\lambda - \sum_f V_{fc}), \quad (1)$$
>
> where ΔV_{fc} is the weight change of a particular feature for the category, α_f is a measure of the salience of the feature, β_c is a measure of the salience of the category that occurs on that trial, λ is the asymptote of associative strength (usually 100), and
>
> $$\sum_f V_{fc}$$
>
> is the sum of the associative strengths of all the features present on that trial for the category that actually occurs. At the same time, the weights connecting each feature that is present to each category that does not occur on that trial are reduced; this represents extinction of the association between the feature and the absent category. Reduction of these weights occurs according to the equation
>
> $$\Delta V_{fc} = -\alpha_f \cdot \gamma_c \cdot \sum_f V_{fc}, \quad (2)$$
>
> where γ_c is a measure of the salience of the category not occurring on that trial and
>
> $$\sum_f V_{fc}$$
>
> now refers to the total associative strength of the symptoms present for the category not occurring on that trial. β_c is usually given a greater value than γ_c, indicating that a category's occurrence is more salient than its omission. (p. 434)

Deriving the Asymptotic Weights

To obtain the asymptotic weights of a network, we begin by noting the equivalence between the Rescorla–Wagner learning rule and the least-mean-squares rule of Widrow and Hoff (1960; cf. Sutton & Barto, 1981). This updating rule implements an iterative algorithm for computing the solution to a set of linear equations, defined by the set of stimulus–response patterns presented to the network. If the input stimulus patterns are linearly independent, then the updating rule will reveal a unique solution. Even if the stimulus patterns are not linearly independent, the network will still converge provided that the learning rate is sufficiently small and that the various stimulus patterns occur with sufficient frequency in the input sequence. The network will converge so as to minimize the sum of the squared error over the stimulus patterns. That is, the equation

$$E = \sum_p \pi_p l_p (\lambda_p - \sum_f V_{fcp})^2 \quad (3)$$

will be minimized, where p is the index for a particular stimulus–response pattern, π_p is the relative frequency of pattern p, l_p is the learning rate associated with pattern p (either β_c or γ_c), λ_p is the desired output for the pattern (either 0 or 100 in Shanks's, 1991, simulations), and $\sum_f V_{fcp}$ is the actual output produced for the pattern, which is equal to the sum of the associative strengths of all present features for the pattern. Each term in Equation 3 is weighted by the learning rate associated with each pattern and by the relative frequency of occurrence of each pattern to reflect the impact a particular pattern has relative to other patterns on weight changes. If the reinforcement learning rate β_c is equal to the nonreinforcement learning rate γ_c, then the l_p term may be omitted from the equation. Similarly, if the patterns occur with uniform frequency, then the π_p term may be omitted from the equation.

Asymptotic Weights for Simulations 1–3

Shanks's (1991) first simulation consisted of two conditions. The contingent condition consisted of one trial in which two input cues, A and B, were associated with Disease 1 and another in which B alone occurred and was not associated with any disease (i.e., it was nonreinforced). The noncontingent condition also had a trial in which two stimuli, C and D, were associated with Disease 2, but in contrast to the contingent condition, on the second trial D was associated with Disease 2 rather than nonreinforced.

The theoretical asymptotic weights of the network can be calculated analytically by minimizing the sum squared error given by Equation 3. To obtain the solution for the asymptotic weights for the contingent condition in Simulation 1, we first derive the appropriate instantiation of Equation 3. Recall that for the contingent condition there are two stimulus patterns. For the first stimulus pattern, the desired output λ_p is 100, and both Cue A and Cue B are present. Hence the contribution to the error for a trial of the first pattern type ($p = 1$) is

$$(\lambda_1 - \sum_f V_{fc1})^2 = [100 - (V_A + V_B)]^2.$$

Similarly, the contribution to the error for a trial of the second pattern type ($p = 2$), where B is present and the desired output is 0, is

$$(\lambda_2 - \sum_f V_{fc2})^2 = [0 - (V_B)]^2$$
$$= V_B^2.$$

We may combine these two terms to calculate the sum of squared errors weighting each term by the learning rate associated with each pattern. Because each pattern type occurs with equal frequency, the sum squared error is

$$E = \sum_p l_p (\lambda_p - \sum_f V_{fcp})^2$$
$$= \beta_c [100 - (V_A + V_B)]^2 + \gamma_c (V_B)^2$$

In general, the minimum value of the sum of squared errors may be obtained by setting the partial derivatives with respect to each weight to 0 and solving the resulting set of equations. In this case, however, the weights that minimize the Equation just mentioned can

(*Appendix continues on next page*)

be obtained by inspection: E will have its lowest value when $V_A + V_B = 100$ and $V_B = 0$. Hence, the asymptotic solution for the contingent condition is $V_A = 100$ and $V_B = 0$.

Similarly, the sum of squared errors for the noncontingent condition is

$$\beta_c[100 - (V_C + V_D)]^2 + \beta_C(100 - V_D)^2.$$

Here, the weights that minimize the equation are $V_C = 0$ and $V_D = 100$.

The asymptotic weights for Simulations 2 and 3 may be derived with the method described above, obtaining for Simulation 2,

Contingent: $V_A = 100$, $V_B = 0$, $V_C = 100$

Noncontingent: $V_D = 0$, $V_E = 100$, $V_F = 0$

and for Simulation 3,

Correlated: $V_B = 100 - V_A$, $V_C = -V_A$

Uncorrelated: $V_E = (\beta_c/(\beta_c + \gamma_c))100 - V_D$, $V_E = V_F$.

Note that whereas Simulations 1 and 2 have unique solutions, which are independent of the initial weights and the learning parameters, there is no unique solution for either the correlated condition or the uncorrelated condition in Simulation 3. In addition to the learning parameters, the only other parameters in the model are the initial weights. In the correlated condition, because the learning parameters drop out of the equations, the only parameters left are the initial weights. The solution for the correlated condition is dependent on the initial weights, and the solution for the uncorrelated condition is dependent both on the learning rate parameters and the initial weights.

Received December 27, 1991
Revision received August 31, 1992
Accepted September 8, 1992 ■

Call for Nominations

The Publications and Communications Board has opened nominations for the editorships of *Behavioral Neuroscience*, the *Journal of Experimental Psychology: General*, and the *Journal of Experimental Psychology: Learning, Memory, and Cognition* for the years 1996–2001. Larry R. Squire, PhD, Earl Hunt, PhD, and Keith Rayner, PhD, respectively, are the incumbent editors. Candidates must be members of APA and should be available to start receiving manuscripts in early 1995 to prepare for issues published in 1996. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. To nominate candidates, prepare a statement of one page or less in support of each candidate.

- For *Behavioral Neuroscience*, submit nominations to J. Bruce Overmier, PhD, Elliott Hall—Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455 or to psyjbo@vx.cis.umn.edu. Other members of the search committee are Norman Adler, PhD, Evelyn Satinoff, PhD, and Richard F. Thompson, PhD.

- For the *Journal of Experimental Psychology: General*, submit nominations to Howard E. Egeth, PhD, Chair, *JEP: General* Search, Department of Psychology, Johns Hopkins University, Charles & 34th Streets, Baltimore, MD 21218, to egeth@jhuvm.bitnet, or to fax number 410-516-4478. Other members of the search committee are Donald S. Blough, PhD, Martha Farah, PhD, and Edward E. Smith, PhD.

- For the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, submit nominations to Donna M. Gelfand, PhD, Dean, Social and Behavioral Science, 205 Osh, University of Utah, Salt Lake City, UT 84112-1102 or to fax number 801-585-5081. Other members of the search committee are Marcia Johnson, PhD, Michael Posner, PhD, Henry L. Roediger III, PhD, and Richard M. Shiffrin, PhD.

First review of nominations will begin December 15, 1993.