

Asymptotic judgment of cause in a relative validity paradigm

A. G. BAKER

McGill University, Montreal, Quebec, Canada

and

FRÉDÉRIC VALLÉE-TOURANGEAU and ROBIN A. MURPHY

University of Hertfordshire, Hatfield, England

We report three experiments in which we tested asymptotic and dynamic predictions of the Rescorla–Wagner (R–W) model and the asymptotic predictions of Cheng's probabilistic contrast model (PCM) concerning judgments of causality when there are two possible causal candidates. We used a paradigm in which the presence of a causal candidate that is highly correlated with an effect influences judgments of a second, moderately correlated or uncorrelated cause. In Experiment 1, which involved a moderate outcome density, judgments of a moderately positive cause were attenuated when it was paired with either a perfect positive or perfect negative cause. This attenuation was robust over a large set of trials but was greater when the strong predictor was positive. In Experiment 2, in which there was a low overall density of outcomes, judgments of a moderately correlated positive cause were elevated when this cause was paired with a perfect negative causal candidate. This elevation was also quite robust over a large set of trials. In Experiment 3, estimates of the strength of a causal candidate that was uncorrelated with the outcome were reduced when it was paired with a perfect cause. The predictions of three theoretical models of causal judgments are considered. Both the R–W model and Cheng's PCM accounted for some but not all aspects of the data. Pearce's model of stimulus generalization accounts for a greater proportion of the data.

In the last decade, research on human judgment of contingency has been motivated by a comparative perspective that stresses an analogous associative learning component in both animal conditioning and human judgments of contingency. When a single predictor variable signals the occurrence of an outcome variable, contingency judgments closely parallel the actual one-way contingency, ΔP , between the predictor and the outcome. ΔP is defined as the difference between the conditional probabilities $p(\text{outcome} | \text{predictor})$ and $p(\text{outcome} | \text{no predictor})$ (Allan, 1980). This parallel has been observed in both operant learning tasks (see, e.g., Wasserman, Elek, Chatlosh, & Baker, 1993) and video games that are analogous to classical conditioning tasks (see, e.g., Dickinson, Shanks, & Evenden, 1984). These results suggest that subjects might derive their judgments from an unbiased consideration of ΔP .

This account of contingency judgment fares poorly in situations where two predictor variables, A and B , signal a common outcome. Usually if predictor A 's contingency is weaker than B 's, then judgments of A 's contingency are reduced. This reduction in estimates of A 's influence is

called *discounting* or *blocking* (we will use these terms interchangeably throughout). In a series of experiments inspired by Kamin's (1969) blocking preparation, Dickinson et al. (1984) demonstrated that contingency estimates of moderately positive predictor A were reliably attenuated by exposing subjects to a strong positive contingency between predictor B and the same outcome in an initial observation stage. In turn, in a series of experiments inspired by the multiple-cue preparation of Wagner, Logan, Haberlandt, and Price (1968), Baker, Mercier, Vallée-Tourangeau, Frank, and Pan (1993) established that estimates of moderately positive or negative contingencies between predictor A and the outcome were influenced by the nature of concurrent strongly positive or negative contingencies between predictor B and the outcome. Similar discounting effects using different causal scenarios have been reported in Chapman and Robbins (1990), Price and Yates (1993, 1995), Shanks (1991), and Vallée-Tourangeau, Baker, and Mercier (1994).

Two general classes of models have been used to attempt to understand discounting. The first class includes associative models arguing that causal judgments are modeled best by rather simple connectionist networks that do not include episodic memories or normative representations of the events. These are process models. That is, they predict both the course of acquisition and the final asymptotic level of judgments. The most common associative model considered has been the Rescorla–Wagner (R–W; 1972) model. The second class of models

This research was supported by a National Sciences and Engineering (Canada) research grant to A.G.B. and an Economic and Social Research Council (U.K.) grant to F.V.-T. Correspondence should be addressed to A. G. Baker, Department of Psychology, McGill University, 1205 Dr. Penfield, Montreal, PQ, H3A 1B1, Canada (e-mail: bakerandy@hebb.psych.mcgill.ca).

includes normative models suggesting that the psychological mechanism behind judgments is best understood if it is postulated that subjects maintain a more or less accurate representation of events and then calculate the contingency or relationship between these events. Some versions of these models include the notion of causal power—that is, that subjects have a mental model of the causal relationship (see, e.g., Cheng, Park, Yarlas, & Holyoak, 1996; Waldmann & Holyoak, 1992). Normative models generally do not predict the course of acquisition of judgments; rather, they typically predict only asymptotic performance. As implied above, associative theories have often been contrasted with a rather naive version of normative models positing that subjects simply calculate the unconditional contingency (ΔP) between cause and effect. These simple normative models do not predict the basic discounting effect. However, more complete versions such as the probabilistic contrast model (PCM; see Cheng et al., 1996) and its power version (PowerPC; see Cheng, 1997).

The present experiments were designed to test the dynamic predictions of the R–W model and contrast its asymptotic predictions with those of the normative PCM model. We chose three results of our earlier research as the forum for this comparison. We will first describe the causal judgment task we use. We will then describe the R–W model and the PCM. Finally, in the General Discussion, we will introduce a more recent associative model that provides a more successful account of our data than either R–W or PCM (i.e., Pearce, 1987).

The Judgment Task

The judgment task designed by Baker et al. (1993) involved a video game in which discrete trials showed a tank traveling toward a field of light-sensitive mines. The subjects could camouflage the tank before it entered the minefield. The camouflage was the target cue (A) for the discounting or blocking treatments in all three experiments described here. The camouflage might make the tank more or less detectable for the mines, and, therefore it could either prevent or precipitate explosions. Thus the camouflage–safety contingency could range between -1 and 1 . In the following description of experimental treatments, the camouflage's contingency was set at $\Delta P = .5$, where $p(\text{safety} | \text{camouflage}) = .75$ and $p(\text{safety} | \text{no camouflage}) = .25$, unless specified otherwise. Subjects were asked to assess the contingency between the camouflage (predictor A) and safety (the outcome). In the present experiments, we will use the notation $A[P(\text{safety} | A) - P(\text{safety} | \text{no } A)]$ to describe the contingencies; thus this contingency would be called $A(.75 - .25)$. Note: the difference between the conditional probabilities in the parentheses equals ΔP . An $A(.75 - .25)$ contingency is called a moderate-density (density here means the expected frequency or probability of outcomes) moderately positive contingency because $\Delta P = .5$ and the outcome density (defined here as the mean of the two conditional probabilities $(.75 - .25)/2$) is $.5$. This would contrast with a high

density moderately positive contingency, $A(1 - .5)$ (outcome density = $.75$), or the low density $A(.5 - 0)$ contingency that we will study in Experiment 2.

The game also involved a second cue (predictor B), a spotter plane that could accompany the tank on some trips through the minefield. The plane was the stimulus (B) that, when perfectly correlated with the outcome, might be expected to cause the subjects to discount the moderately positive camouflage. Since the plane could be present or absent on either safety or explosion trials, the plane–safety contingency could also range between -1 and 1 . In one discounting treatment (Baker et al., 1993, Experiment 1), the spotter plane appeared on all safety trials (i.e., $p(\text{safety} | \text{plane}) = 1$), but never on unsafe trials $p(\text{safety} | \text{no plane}) = 0$; contingency name: $B(1 - 0)$. As noted, the contingency for the camouflage was $A(.75 - .25)$, so, using the notation (A contingency/ B contingency), this treatment would be called $A(.75 - .25)/B(1 - 0)$. (Baker et al., 1993, used the notation $(\Delta P_A/\Delta P_B)$; thus they called this treatment $.5/1$.) In a second, control, treatment, $A(.75 - .25)/B(.5 - .5)$, the plane was not correlated with safety (i.e., $p(\text{safety} | \text{plane}) = .5$ and $p(\text{safety} | \text{no plane}) = .5$). As in Wagner et al. (1968), human judgments of the effectiveness of the camouflage were reliably lower when, in the discounting treatment, the plane signaled all the occurrences of safety, even though in both conditions the camouflage contingency equaled $.5$.

The R–W Model

The fundamental discounting or blocking effect is well modeled by the R–W model, which can be conceptualized as a single-layer adaptive network composed of three input units and one output unit. Two of the input units represent the two predictor variables or causal cues (A and B), and the third (X) represents the context in which the cues are presented. The activation of the output unit is equal to its net input. On a given training trial, the changes in the connection weight or associative strength of input cue (i) are calculated on the basis of the following rule:

$$\Delta V_i = \alpha_i \beta (\lambda - \sum V_k)$$

Changes in associative strength of cue i are thus a function of the discrepancy between the sum of the associative strengths of all presently active input units (i.e., of the cues that are present on that trial; $\sum V_k$) and the target output activation, or asymptote for learning (λ). λ is positive when the outcome variable is present on that trial and 0 when it is absent; α_i and β are learning rate parameters determined by the saliences of cue i and the outcome, respectively. Importantly, the linear operator $\alpha_i \beta (\lambda - \sum V_k)$ becomes negative when the sum of the associative strengths of all cues present is greater than the asymptote for learning (when $\sum V_k > \lambda$). When this happens, the model will generate a decline in the associative strengths of any cues that are present. Because of this property, the sum of the weights for the three connections will usually approach and tend to be limited by λ . Thus, in many situations, the stronger the connection

strength of one input cue, the weaker will be that of others. It is, thus, a characteristic of this model that individual cues compete for associative strength.

The model accounts for subjects' ability to judge accurately the contingency between a single cue and the outcome, because A 's associative strength V_A correlates well with ΔP . This comes about because A competes with the context X for associative strength. When A 's correlation with the outcome is strongly positive, its associative strength approaches λ , and that of the context approaches zero. When the cue is uncorrelated with the outcome, its associative strength will be zero and the context's will be positive. When A is negatively correlated with the outcome, V_A will become negative through a two-step process. The context X will first develop positive associative strength. As this happens, the linear operator on the majority of AX trials that are not paired with the outcome will become negative, and this will force V_A to become negative. This occurs because on no-outcome trials, the learning asymptote λ equals 0 and $\sum V_k$ [i.e., $(V_X + V_A)$] is greater than 0; hence the linear operator $(\lambda - \sum V_k)$ is negative. This forces V_A to be negative, and it can not recover because A is never or rarely paired with the outcome.

Of particular interest are the model's predictions regarding the amount of associative strength for predictor A in the discounting treatment, $A(.75 - .25)/B(1 - 0)$, and control treatment, $A(.75 - .25)/B(.5 - .5)$, from Baker et al. (1993). The model's predictions for the associative strength of A , B , and X for the three experiments are shown in Figure 1. The left panels (Experiment 1) show the relevant values. When the network is presented with the event contingencies of Treatment $A(.75 - .25)/B(.5 - .5)$, the associative strength, or the connection weight, of predictor A (the camouflage) approximates A 's actual contingency just as it would in the single-cue experiment described above (see the open circles in the top left panel). However, in discounting Treatment $A(.75 - .25)/B(1 - 0)$, A and B are positive predictors whose combined associative strength is constrained by λ (see the triangles). A loses associative strength on some trials because, on these trials, it is not paired with the outcome, but B never loses strength because it is always paired with the outcome. Therefore B 's associative strength will approach λ and this will eventually prevent A from regaining those losses in associative strength. Thus at asymptote the associative strength of B will approach one and that of A and the context X will be near zero.

The R-W network also makes the rather more counterintuitive prediction that discounting of A will occur when B is a perfectly negative predictor of the outcome (i.e., $\Delta P = [p(\text{outcome} | B) - p(\text{outcome} | \text{not } B)] = (0 - 1) = -1$; treatment $B[0 - 1]$). This negative discounting occurs in several stages in which the negative predictor B develops strong negative associative strength, and this modulates the competition between the target cue A and the context X for positive associative strength. This competition is ultimately won by X . X blocks A partly because it is paired with the outcome alone on occasion

and is paired with it more often than is A . But in addition, X , rather paradoxically, acquires extra associative strength on the BX extinction trials (i.e., when $BX \rightarrow$ no outcome). This happens because when B becomes strongly inhibitory, the magnitude of its connection weight (V_B) will approximately equal the complement of the sum of the associative weights of A and the context X [$V_B \cong -(V_A + V_X)$]. That is, its inhibitory properties will approximately equal the sum of their excitatory properties. This is another equilibrium forced by the linear operator. This implies that $V_B < (-V_X)$ and on $BX \rightarrow$ no-outcome trials, $\sum V_k = (V_B + V_X)$ will be a negative quantity. Because $\lambda = 0$ on no-outcome trials, the operator $(\lambda - \sum V_k)$ will be positive. Thus, X will acquire positive associative strength on these trials. This "extra" connection weight will ultimately help X win the competition with A for positive associative strength. As the panel from Figure 1 showing B 's associative strength shows, the ability of X to block associative strength to A is modulated by negative associative strength acquired by B . Although the model predicts equal discounting at asymptote, this two-stage process implies that discounting or blocking will occur more slowly in the $A(.75 - .25)/B(0 - 1)$ treatment than in $A(.75 - .25)/B(1 - 0)$, in which positive associative strength to B directly blocks A .

The Probabilistic Contrast Model

Since A 's ΔP was kept constant at .5 in the discounting treatments, a simple ΔP model arguing that reasoners use this statistically normative rule to inform their judgments cannot explain why predictor A is discounted in some treatments. However, Cheng and Holyoak's (1995) PCM (see also Cheng & Novick, 1992; Melz, Cheng, Holyoak, & Waldmann, 1993; Waldmann & Holyoak, 1992) postulates that reasoners assess A 's contingency by calculating ΔP (or in their terminology, "probabilistic contrasts") for A conditional on the presence and on the absence of B . The conditional ΔP s for the three experiments we report here are shown in Table 1. Contingencies $A(.75 - .25)/B(1 - 0)$, $A(.75 - .25)/B(.5 - .5)$, and $A(.75 - .25)/B(0 - 1)$ are shown under the heading "Experiment 1." Thus, in Treatment $A(.75 - .25)/B(1 - 0)$, A 's contingency in the presence of B ($\Delta P_{A|B}$) is 0, since the probability of an outcome on those trials equals 1 regardless of the presence or the absence of A ; that is, $p(\text{outcome} | A \& B) = p(\text{outcome} | \text{no } A \& B) = 1$. In addition, in the absence of B , A 's contingency ($\Delta P_{A|no B}$) is also 0, since the probability of an outcome on those trials equals 0; that is, $p(\text{outcome} | A \& \text{no } B) = p(\text{outcome} | \text{no } A \& \text{no } B) = 0$. These conditional ΔP s indicate that A is not a true cause of the outcome. Put another way, once the subject knows about the presence of B , A provides no extra information concerning the outcome. Thus the subject will attribute no causal importance to it. Conditional ΔP s in Treatment $A(.75 - .25)/B(0 - 1)$ also reveal that A is a spurious cause of the outcome. The PCM thus predicts discounting of predictor A in Treatments $A(.75 - .25)/B(1 - 0)$ and $A(.75 - .25)/B(0 - 1)$.

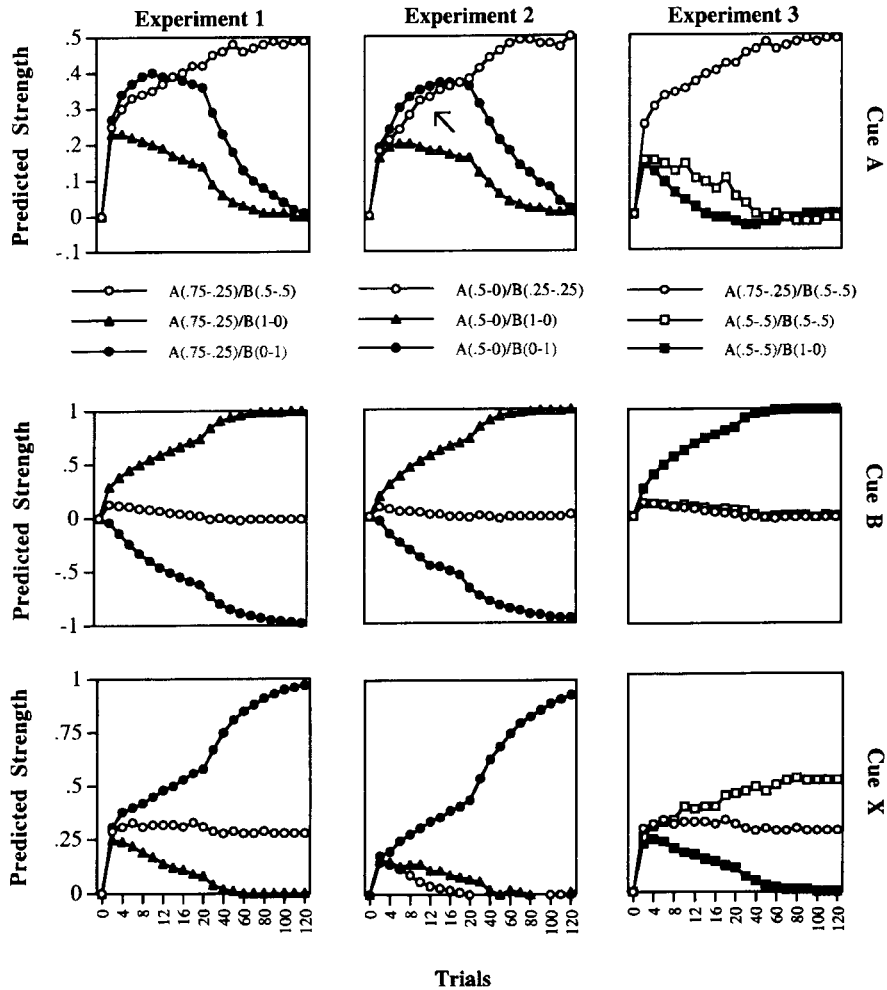


Figure 1. Predicted associative strengths from the Rescorla-Wagner model for the discounted stimulus (the camouflage, Cue *A*) for the three treatments in each of the three experiments are shown in the top three panels. Associative strengths for the blocking stimulus (plane, Cue *A*) and the context are shown in the middle three panels and the bottom three panels, respectively. For all simulations, $\alpha = .35$ for all cues, $\beta_1 = \beta_0 = 1$, and $\lambda = 1$ on trials where the outcome was present and = 0 otherwise; the mean connection weight was calculated over 100 iterations.

Both the PCM and R-W model predict equivalent discounting for both positive and negative *B*s at asymptote. However, because PCM is a competence model, it only makes predictions at asymptote. Conversely, the R-W model is a process model, so it makes dynamic predictions throughout the course of learning, culminating with its asymptotic predictions, which are similar to those of the PCM. Clearly, if the R-W model and PCM are to be compared, it is important to include enough trials to approach asymptote. As shown above, the R-W model predicts faster acquisition of blocking in Treatment $A(.75 - .25)/B(1 - 0)$ than in $A(.75 - .25)/B(0 - 1)$. Furthermore, in our previous research, we discovered two situations in which people's judgments were strongly inconsistent with asymptotic predictions of both the R-W model and the PCM, but were consistent with dynamic pre-

asymptotic predictions of the R-W model. These phenomena will be described and investigated in Experiments 2 and 3.

Our previous experiments have all used 40 training trials. In order to test the dynamic predictions of the R-W model and to ensure that the subjects had reached asymptote in order to justify comparisons of the R-W model and PCM, the present experiments used 120 trials for each treatment. The subjects were asked for contingency judgments at 20, 40, and 120 trials. We chose to use a maximum of 120 trials because this is at least six times the number of trials that it took for estimates to approach an apparent asymptote in our previous research (e.g., Baker, Berbrier, & Vallée-Tourangeau, 1989). It is also the maximum number of trials that permit us to expose the subjects to three different contingencies in a 40-50 min session.

Table 1
Conditional ΔP for the Camouflage (A) Based on the Presence and Absence of the Plane (B), Experiments 1–3

Treatment	Universal Set ΔP_A	Focal Set					
		Plane Present			Plane Absent		
		$p(O A\&B)$	$p(O no\ A\&B)$	$\Delta P_{A B}$	$p(O A\&no\ B)$	$p(O no\ A\&no\ B)$	$\Delta P_A no\ B$
Experiment 1							
$A(.75 - .25)/B(.5 - .5)$.5	.75	.25	.5	.75	.25	.5
$A(.75 - .25)/B(1 - .5)$.5	1	1	0	0	0	0
$A(.75 - .25)/B(0 - 1)$.5	0	0	0	1	1	0
Experiment 2							
$A(.5 - 0)/B(.25 - .25)$.5	.5	0	.5	.5	0	.5
$A(.5 - 0)/B(1 - 0)$.5	1	Undefined	—	0	0	0
$A(.5 - 0)/B(0 - 1)$.5	0	0	0	1	Undefined	—
Experiment 3							
$A(.75 - .25)/B(.5 - .5)$.5	.75	.25	.5	.75	.25	.5
$A(.5 - .5)/B(.5 - .5)$	0	.5	.5	0	.5	.5	0
$A(.5 - .5)/B(1 - 0)$	0	1	1	0	0	0	0

Note—O, outcome.

EXPERIMENT 1

Experiment 1 explored the dynamic nature of the discounting of a moderate-density, moderately positive predictor (A) in the presence of a second predictor (B) that is either a perfectly positive predictor or a perfectly negative predictor of the outcome. As noted, the R–W model predicts that in both these treatments, $A(.75 - .25)/B(1 - 0)$ and $A(.75 - .25)/B(0 - 1)$, the magnitude of discounting should increase over time. But, although the model predicts equivalent asymptotic discounting in both treatments, the two-stage process predicts that discounting will occur more slowly in $A(.75 - .25)/B(0 - 1)$ than in $A(.75 - .25)/B(1 - 0)$. This prediction is clearly shown in Figure 1. The PCM also predicts equivalent asymptotic discounting but makes no predictions concerning the rate of acquisition of discounting.

The nominal contingency tables and conditional contingency tables for this experiment and the second and third experiments are shown in Table 2. Because the subjects choose when to camouflage the tank, different subjects have different numbers of each type of trial and slightly different contingencies. The mean group contingencies very closely mapped onto the nominal contingencies.

Method

Subjects. Twenty-two undergraduates from McGill University volunteered for this experiment.

Task characteristics. Subjects monitored different event contingencies through a video game involving light-sensitive mines, camouflage, and a spotter plane, as described earlier. The task was divided into three “games” corresponding to three experimental treatments. A trial in each of these games consisted of a tank traversing the video screen from right to left to enter the minefield, and subjects could choose to camouflage the tank or not. In order to camouflage the tank, they aimed a “paint-o-ray” that sat immobile at the bottom center of the screen and fired (by pressing the space bar) a paint “bullet” that, upon hitting the tank, changed the tank’s color from blue to green. There were two tank speeds (3.3 cm/sec and 6.0 cm/sec) selected randomly for every trial. A hit was not dif-

ficult to achieve. Pilot subjects instructed solely to hit the tank could do so over 80% of the time. However, the instructions encouraged subjects to hit the tank on some occasions but to let it go by on others. In Experiment 1, subjects hit the tank on an average of 68% of the trials and a similar number of times for each treatment. The spotter plane was programmed to appear above the tank with a certain probability after the tank passed the sight of the paint-o-ray—that is, after the subjects chose to camouflage the tank or not. The aim of the game was to evaluate the effectiveness of the camouflage and the spotter plane in “protecting” the tanks during the game.

Design. Three treatments were designed that corresponded to the three games played by the subjects. In all three treatments the camouflage–safety contingency was set at .5. The probability of a safety trial given that the tank was camouflaged, $p(\text{safety}|\text{camouflage})$, equaled .75, and the probability of a safety trial given that the tank was not camouflaged, $p(\text{safety}|\text{no camouflage})$, equaled .25. In Treatment $A(.75 - .25)/B(.5 - .5)$, the plane–safety contingency was 0: $p(\text{safety}|\text{plane}) = p(\text{safety}|\text{no plane}) = .5$. In Treatment $A(.75 - .25)/B(1 - 0)$, the plane–safety contingency was 1: $p(\text{safety}|\text{plane}) = 1$, $p(\text{safety}|\text{no plane}) = 0$. Finally, in Treatment $A(.75 - .25)/B(0 - 1)$, the plane–safety contingency was –1: $p(\text{safety}|\text{plane}) = 0$, $p(\text{safety}|\text{no plane}) = 1$.

Procedure. Subjects were tested individually. They read the description of the scenario and the game instructions from the screen. Before the start of the first game, subjects practiced camouflaging the tank for 10 trials.

The order of the treatments was randomized for each subject. Each game lasted 120 trials. Subjects made effectiveness ratings of the camouflage and the plane after 20, 40, and 120 trials. Ratings were made using a –100–100 scale. The subjects were informed that negative estimates indicated that the camouflage or the plane reduced the safety of the tank, whereas positive estimates indicated that the camouflage or the plane increased the safety of the tank (i.e., reduced the likelihood that it would explode). They were also informed that a rating of zero represented no relationship between cause and effect. The camouflage ratings screen (which used a blue background) always came before the plane ratings screen (which used a red background) at each of the three rating intervals.

Results

Camouflage ratings. The mean effectiveness ratings, and their standard errors, for the camouflage are reported in Table 3 and in the left panel of Figure 2. Inspection of

Table 2
Unconditional and Conditional Probabilities of the Occurrence of the Outcome in the Presence and Absence of the Camouflage (*A*) and the Plane (*B*), Experiments 1-3

			Experiment 1																																																					
$A(.75 - .25)/B(.5 - .5)$			$A(.75 - .25)/B(1 - 0)$			$A(.75 - .25)/B(0 - 1)$																																																		
<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>.75</td><td>.25</td><td>.5</td></tr> <tr><td><i>-B</i></td><td>.75</td><td>.25</td><td>.5</td></tr> <tr><td></td><td>.75</td><td>.25</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	.75	.25	.5	<i>-B</i>	.75	.25	.5		.75	.25		<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>1</td><td>1</td><td>1</td></tr> <tr><td><i>-B</i></td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>.75</td><td>.25</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	1	1	1	<i>-B</i>	0	0	0		.75	.25		<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>0</td><td>0</td><td>0</td></tr> <tr><td><i>-B</i></td><td>1</td><td>1</td><td>1</td></tr> <tr><td></td><td>.75</td><td>.25</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	0	0	0	<i>-B</i>	1	1	1		.75	.25	
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	.75	.25	.5																																																					
<i>-B</i>	.75	.25	.5																																																					
	.75	.25																																																						
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	1	1	1																																																					
<i>-B</i>	0	0	0																																																					
	.75	.25																																																						
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	0	0	0																																																					
<i>-B</i>	1	1	1																																																					
	.75	.25																																																						
			Experiment 2																																																					
$A(.5 - 0)/B(.25 - .25)$			$A(.5 - .25)/B(1 - 0)$			$A(.5 - 0)/B(0 - 1)$																																																		
<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>.5</td><td>0</td><td>.25</td></tr> <tr><td><i>-B</i></td><td>.5</td><td>0</td><td>.25</td></tr> <tr><td></td><td>.5</td><td>0</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	.5	0	.25	<i>-B</i>	.5	0	.25		.5	0		<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>1</td><td>-</td><td>1</td></tr> <tr><td><i>-B</i></td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>0.5</td><td>0</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	1	-	1	<i>-B</i>	0	0	0		0.5	0		<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>0</td><td>0</td><td>0</td></tr> <tr><td><i>-B</i></td><td>1</td><td>-</td><td>1</td></tr> <tr><td></td><td>0.5</td><td>0</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	0	0	0	<i>-B</i>	1	-	1		0.5	0	
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	.5	0	.25																																																					
<i>-B</i>	.5	0	.25																																																					
	.5	0																																																						
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	1	-	1																																																					
<i>-B</i>	0	0	0																																																					
	0.5	0																																																						
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	0	0	0																																																					
<i>-B</i>	1	-	1																																																					
	0.5	0																																																						
			Experiment 3																																																					
$A(.75 - .25)/B(.5 - .5)$			$A(.5 - .5)/B(.5 - .5)$			$A(.5 - .5)/B(1 - 0)$																																																		
<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>.75</td><td>.25</td><td>.5</td></tr> <tr><td><i>-B</i></td><td>.75</td><td>.25</td><td>.5</td></tr> <tr><td></td><td>.75</td><td>.25</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	.75	.25	.5	<i>-B</i>	.75	.25	.5		.75	.25		<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>.5</td><td>.5</td><td>.5</td></tr> <tr><td><i>-B</i></td><td>.5</td><td>.5</td><td>.5</td></tr> <tr><td></td><td>.5</td><td>.5</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	.5	.5	.5	<i>-B</i>	.5	.5	.5		.5	.5		<table border="1"> <tr><td></td><td><i>A</i></td><td><i>-A</i></td><td></td></tr> <tr><td><i>B</i></td><td>1</td><td>1</td><td>1</td></tr> <tr><td><i>-B</i></td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>.5</td><td>.5</td><td></td></tr> </table>				<i>A</i>	<i>-A</i>		<i>B</i>	1	1	1	<i>-B</i>	0	0	0		.5	.5	
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	.75	.25	.5																																																					
<i>-B</i>	.75	.25	.5																																																					
	.75	.25																																																						
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	.5	.5	.5																																																					
<i>-B</i>	.5	.5	.5																																																					
	.5	.5																																																						
	<i>A</i>	<i>-A</i>																																																						
<i>B</i>	1	1	1																																																					
<i>-B</i>	0	0	0																																																					
	.5	.5																																																						

Note—The individual subcells indicate the conditional probability of outcome based on each conjunction of the presence and absence of the cues [e.g., cell *A*, *-B* represents $P(\text{outcome} | \text{camouflage and no plane})$]. The row and cell marginal values represent the overall conditional probabilities based on the presence and absence of the camouflage or the plane [e.g., the marginal value of column *-A* represents $P(\text{outcome} | \text{no camouflage})$].

the data indicates that each of the three treatments generated a different level of estimates. The control treatment, $A(.75 - .25)/B(.5 - .5)$, generated estimates that approximated $\Delta P_{\text{camouflage}} \times 100$ and appeared to be constant over the full 120 trials. The estimates of both Treatments $A(.75 - .25)/B(1 - 0)$ and $A(.75 - .25)/B(0 - 1)$ appeared to decline over trials, and those of $A(.75 - .25)/B(1 - 0)$ appeared to be lower than those of $A(.75 - .25)/B(0 - 1)$. A two-factor repeated measures analysis of variance (ANOVA) on these data supports some of the above observations. It revealed a reliable main effect of treatment [$F(2,42) = 18.0$] as well as a reliable main effect of trials [$F(2,42) = 5.44$], but the interaction was not reliable [$F(4,84) = 1.48$] (we use a 5% rejection rate throughout).

The absence of a reliable interaction in the preceding analysis does not support the notion that discounting developed over trials and that it developed more rapidly in the $A(.75 - .25)/B(1 - 0)$ treatment. However, in a design that postulates learning curves that develop at different rates, the postulated effect is confounded between the interaction and both main effects of a conventional split-plot ANOVA. We further investigated both the interaction and the main effects using post hoc ANOVAs. These tests were corrected for chance using the Bonferroni method. In an analysis comparing the two discounting treatments,

$A(.75 - .25)/B(1 - 0)$ and $A(.75 - .25)/B(0 - 1)$, the main effects of treatments and trials were each reliable [$F(1,21) = 9.95$; $F(2,42) = 9.70$], whereas the interaction was not [$F(2,42) = .061$]. This analysis indicates that estimates decreased over trials and that the estimates in the $A(.75 - .25)/B(1 - 0)$ treatment were lower than those for the $A(.75 - .25)/B(0 - 1)$ treatment. This difference was fairly constant throughout. The decline in estimates occurred mostly between 20 and 40 trials; the difference in estimates between 40 and 120 trials was not reliable [$F(1,21) = 1.49$]. There was no reliable change in estimates over trials for the nondiscounting control treatment, $A(.75 - .25)/B(.5 - .5)$ [$F(2,42) = .013$]. Although treatment $A(.75 - .25)/B(1 - 0)$ differed from the control treatment, $A(.75 - .25)/B(.5 - .5)$, after 20, 40, and 120 trials, Treatment $A(.75 - .25)/B(0 - 1)$ only differed reliably from Treatment $A(.75 - .25)/B(.5 - .5)$ after 40 and 120 trials [$F(1,21) = 6.38$]. Thus the statistical analysis supports the argument that discounting seems to develop over the first 40 trials but does not support the argument that discounting will ultimately be equal following the $A(.75 - .25)/B(1 - 0)$ and $A(.75 - .25)/B(0 - 1)$ treatments.

Plane ratings. The mean effectiveness ratings for the plane are reported in Table 4. Subjects easily recognized the nature of the contingencies, with terminal estimates

Table 3
Mean Effectiveness Ratings and Standard Errors for the Camouflage

Treatment	Trials					
	20		40		120	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1						
<i>A</i> (.75 - .25)/ <i>B</i> (.5 - .5)	49.3	8.6	48.2	8.6	48.8	8.4
<i>A</i> (.75 - .25)/ <i>B</i> (0 - 1)	43.0	9.8	26.8	8.5	18.6	7.7
<i>A</i> (.75 - .25)/ <i>B</i> (1 - 0)	8.4	10.9	-12.5	10.1	-18.2	10.2
Experiment 2						
<i>A</i> (.5 - 0)/ <i>B</i> (.25 - .25)	50.3	7.3	45.8	8.4	29.6	8.1
<i>A</i> (.5 - 0)/ <i>B</i> (0 - 1)	66.6	9.2	71.6	9.3	65.3	12.7
<i>A</i> (.5 - 0)/ <i>B</i> (1 - 0)	-21.9	14.0	-33.4	12.3	-21.2	11.7
<i>A</i> (.75 - .25)/ <i>B</i> (.5 - .5)	40.4	8.6	54.9	8.5	54.4	8.7
<i>A</i> (.5 - .5)/ <i>B</i> (.5 - .5)	16.9	10.6	7.7	11.6	-7.6	10.3
<i>A</i> (.5 - .5)/ <i>B</i> (1 - 0)	-22.1	12.5	-25.9	11.4	-20.7	13.0

of 97.8 for the perfectly positive contingency, -98.8 for the perfectly negative contingency, and -0.2 for the zero contingency. A two-factor repeated measures ANOVA on these ratings yielded a reliable main effect of treatment [$F(2,42) = 328$], but neither the trials main effect ($F < 1$) nor the interaction was reliable [$F(4,84) = 1.18$].

Discussion

Experiment 1 provides some support for the predictions of the R-W model that discounting will increase over trials and that discounting will develop earlier in Treatment *A*(.75 - .25)/*B*(1 - 0) than in *A*(.75 - .25)/*B*(0 - 1). However, the model predicts that at asymptote, the magnitude of the discounting should be equivalent in both treatments (see Figure 1, left panel). Yet, this was clearly not the case, since judgments of predictor *A* in treatment *A*(.75 - .25)/*B*(1 - 0) were more negative than in treatment *A*(.75 - .25)/*B*(0 - 1).

These data also pose problems for various versions of the PCM. The conditional contrasts for the camouflage are shown in Table 1. These contrasts are consistent with the moderately positive judgments in treatment *A*(.75 - .25)/*B*(.5 - .5), but do not explain either (1) the increase in

discounting over time, or (2) the difference in the magnitude of discounting in Treatments *A*(.75 - .25)/*B*(1 - 0) and *A*(.75 - .25)/*B*(0 - 1) at asymptote.

When there are facilitatory alternative causes, PCM argues that the most informative contrasts to determine the effectiveness of a facilitatory or excitatory cause are those in the absence of any alternative causes. Furthermore, to determine whether a cause is ineffective, the relevant zero contrast must not be at or near the behavioral ceiling. Table 1 shows that in the *A*(.75 - .25)/*B*(1 - 0) contingency, the contrast for the camouflage in the absence of the plane alternative cause consists of conditional probabilities of zero, so it is clear that the camouflage has no causal power independent of the plane and thus should be overshadowed. However, in the case of the *A*(.75 - .25)/*B*(0 - 1) contingency, the zero contrast in the absence of the plane consists of two conditional probabilities of 1, such a contrast is not informative concerning the possible positive casual power of the camouflage. Thus, it could be argued that the greater degree of discounting found at asymptote in the *A*(.75 - .25)/*B*(1 - 0) treatment is predicted by PCM. However, the analysis has its problems because the plane in *A*(.75 -

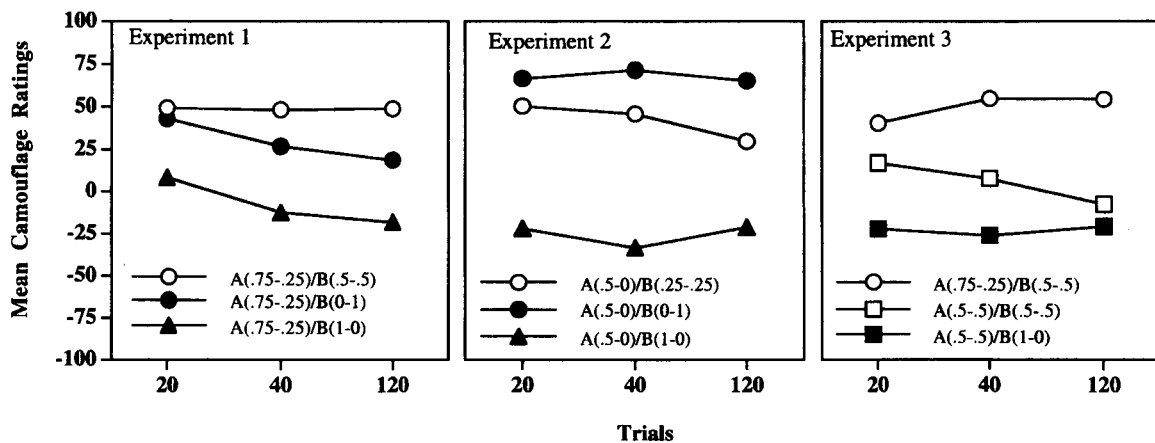


Figure 2. Mean effectiveness ratings for the camouflage in Experiments 1, 2, and 3.

Table 4
Mean Effectiveness Ratings and Standard Errors for the Spotter Plane

Treatment	Trials					
	20		40		120	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Experiment 1						
$A(.75 - .25)/B(.5 - .5)$	4.8	12.6	-3.7	11.5	-0.2	6.3
$A(.75 - .25)/B(0 - 1)$	-95.4	3.1	-95.5	2.8	-98.8	1.2
$A(.75 - .25)/B(1 - 0)$	84.2	6.7	95.2	2.6	97.8	1.3
Experiment 2						
$A(.5 - 0)/B(.25 - .25)$	5.3	11.7	-3.2	10.6	0.4	9.2
$A(.5 - 0)/B(0 - 1)$	-91.1	5.4	-99.2	0.5	-99.5	0.5
$A(.5 - 0)/B(1 - 0)$	87.9	6.0	86.3	7.0	98.4	1.3
Experiment 3						
$A(.75 - .25)/B(.5 - .5)$	10.8	14.1	-10.4	12.4	-4.9	14.3
$A(.5 - .5)/B(.5 - .5)$	-7.8	13.6	12.5	9.8	0.4	11.4
$A(.5 - .5)/B(1 - 0)$	81.9	9.7	92.7	5.9	96.8	3.1

$.25)/B(0 - 1)$ is a negative alternative cause. Thus, to extrapolate the arguments presented in Cheng and Holyoak (1995), the most informative focal set for the camouflage is conditionalized on the presence of the inhibitory plane, not its absence. Neither Cheng and Holyoak nor Cheng et al. (1996) have acknowledged negative alternative conditionalizing cues, but the logic of their analysis of excitatory and inhibitory target causes must extend to alternative causes that are themselves either excitatory or inhibitory. If the PCM considers the presence of preventive causes as defining the focal set for a target cause, it must predict strong discounting in the $A(.75 - .25)/B(0 - 1)$ contingency. In the General Discussion, we will outline Pearce's (1987) model, which accounts more successfully for the present results.

EXPERIMENT 2

Although both the PCM and the R-W model predict the basic positive and negative blocking effects at asymptote, the heuristic value of the R-W simulations is illustrated by the discovery that under some conditions, opposite polarity discounting could be retarded and even reversed. This happened when A was involved in the low-density, moderately positive contingency with the outcome described earlier [i.e., $A(.5 - 0)$]. In this contingency, the outcome occurs only in the presence of A [$p(\text{outcome} | A) = .5$ and $p(\text{outcome} | \text{no } A) = 0$]. According to the model (see Figure 1, middle column), in the negative discounting treatment, $A(.5 - 0)/B(0 - 1)$, the context will still ultimately overshadow judgments of A just as it would in the moderate-density discounting contingency [$A(.75 - .25)/B(0 - 1)$]. However, discounting is retarded and even briefly reversed relative to the $A(.5 - 0)/B(.25 - .25)$ discounting control (note section marked with an arrow in Figure 1). There are two factors responsible for this effect. First, in an $A(.5 - 0)$ treatment, the context is never paired with the outcome on its own. This slows acquisition to the context because, in other contingencies, on context-alone outcome trials, the context can acquire associative

strength without competition. Furthermore, because there are no $AX \rightarrow$ outcome trials in this contingency, the only mechanism for X to block A is the (paradoxical) excitation acquired on $BX \rightarrow$ no outcome trials that was described earlier for the $A(.75 - .25)/B(0 - 1)$ contingency. Second, $A(.5 - 0)$ treatments have a lower outcome density than other contingencies sharing the same ΔP . Thus, because outcome density is lower in $A(.5 - 0)$ than in $A(.75 - .25)$ contingencies, acquisition of associative strength by the context will be slower.

Retarding the context's rate of acquisition of associative strength reduces the discounting effect on predictor A and predicts the transient situation in which judgments of A in the $A(.5 - 0)/B(0 - 1)$ contingency are higher than those in the control $A(.5 - 0)/B(.25 - .25)$ contingency. However, these differences are transient and, at asymptote, judgments of A following both the low-density discounting treatment [$A(.5 - 0)/B(0 - 1)$] and the moderate-density discounting procedure [$A(.75 - .25)/B(0 - 1)$] should be very similar and should both exhibit discounting.

Experiment 5 of Baker et al. (1993) provides evidence that partly supports this prediction. Predictor A 's contingency was not discounted in Treatment $A(.5 - 0)/B(0 - 1)$, and judgments of A were reliably more positive in that treatment than in Treatment $A(.5 - 0)/B(.25 - .25)$. However, although the model predicts that this reversal of blocking is only transient, judgments of A were not reduced in the presence of B 's perfectly negative contingency at any point during the task. However, because our task involved only 40 trials and the function that maps the number of input vectors for an associative network onto the number of trials experienced by human subjects cannot be determined precisely, it is possible that human judgments at Trial 40 do not correspond to the network's asymptotic prediction. Thus, with a much longer judgment task, predictor A 's contingency may come to be discounted rather than enhanced.

In this experiment, the task scenario from Experiment 1 was again used. The camouflage acting as predictor A was again moderately correlated with safety ($\Delta P = .5$) but

was involved in the low-density, moderately positive $A(.5 - 0)$ contingency with the outcome. A spotter plane acted as predictor B and was related to the outcome by contingencies of $B(1 - 0)$, $B(.25 - .25)$, and $B(0 - 1)$ ($\Delta P = 1, 0$, and -1). Each subject was exposed to all three treatments [$A(.5 - 0)/B(1 - 0)$, $A(.5 - 0)/B(.25 - .25)$, and $A(.5 - 0)/B(0 - 1)$] and made estimates of the causal status of the plane and camouflage after 20, 40, and 120 trials.

Method

Subjects. Eighteen undergraduates from McGill University volunteered for this experiment.

Design and Procedure. The camouflage contingency in all three games was set at .5, with $p(\text{safety} | \text{camouflage}) = .5$ and $p(\text{safety} | \text{no camouflage}) = 0$. In Treatment $A(.5 - 0)/B(.25 - .25)$, the plane-safety contingency was 0, with $p(\text{safety} | \text{plane}) = p(\text{safety} | \text{no plane}) = .25$. In treatment $A(.5 - 0)/B(1 - 0)$, the plane-safety contingency was perfect at 1, with $p(\text{safety} | \text{plane}) = 1$ and $p(\text{safety} | \text{no plane}) = 0$. And in Treatment $A(.5 - 0)/B(0 - 1)$, the plane-safety contingency was -1 , with $p(\text{safety} | \text{plane}) = 0$ and $p(\text{safety} | \text{no plane}) = 1$. The nominal contingency tables and conditional contingency tables for Experiment 1 are shown in Table 2.

The same procedure as in Experiment 1 was used. Each treatment lasted 120 trials and order of presentation was randomized for each subject. Subjects camouflaged the tank on 73% of the trials. The effectiveness of the camouflage and of the plane was rated after 20, 40, and 120 trials.

Results

Camouflage ratings. The mean effectiveness ratings and their standard errors for the camouflage contingencies are reported in Table 3 and in Figure 2. The ratings for the $A(.5 - 0)/B(.25 - .25)$ treatment were moderately positive throughout the task, although the terminal rating was lower than the rating after 40 trials. In contrast, the same moderate contingency in the presence of a perfectly positive plane-safety contingency in Treatment $A(.5 - 0)/B(1 - 0)$ was clearly discounted. Judgments were negative over the entire task. In turn, the camouflage contingency was not discounted in the presence of a perfectly negative plane-safety contingency, $A(.5 - 0)/B(1 - 0)$, and, in fact, judgments were more positive than in the $A(.5 - 0)/B(.25 - .25)$ treatment. Both results closely parallel the earlier finding that used only 40 trials, and the latter is clearly inconsistent with the R-W prediction that the elevated judgments in the negative blocking group, $A(.5 - 0)/B(1 - 0)$, should be transient. A two-factor repeated measures ANOVA revealed a reliable main effect of treatment [$F(2,34) = 27.6$], but neither the main effect of trials ($F < 1$) nor the interaction [$F(4,68) = 1.49$] was reliable. This experiment consisted of a control treatment [$A(.5 - 0)/B(.25 - .25)$] and two experimental treatments [$A(.5 - 0)/B(1 - 0)$ and $A(.5 - 0)/B(0 - 1)$]. In order to investigate the reliable main effect found in the preceding ANOVA, we carried out two subsequent ANOVAs. One ANOVA compared the control treatment with each experimental treatment. The main

effect for contingency of these analyses represents contrasts within the reliable main effect. According to conventional wisdom, investigation of the interactions in these analyses was not justified because the interaction in the overall analysis was not reliable. Nevertheless, because the R-W model predicts an interaction (the differential acquisition of discounting in the two experimental groups), these comparisons might be considered preplanned. Furthermore, accepting the claim that the model does not account for our results involved accepting the null hypothesis that the level of estimates did not vary (was not transient), and thus we needed a powerful analysis to justify this claim. In order to control for the inflation in Type I error rate, we used the Bonferroni technique to control for the number of post hoc ANOVAs. Because the present experiment involved two post hoc ANOVAs, we used a 2.5% rejection criterion.

These post hoc analyses indicated that the estimates for the $A(.5 - 0)/B(0 - 1)$ treatment were reliably higher than those of the control [$A(.5 - 0)/B(.25 - .25)$] treatment [$F(1,17) = 8.38$] and that those of Treatment $A(.5 - 0)/B(1 - 0)$ were reliably lower [$F(1,17) = 33.2$]. Thus the present data replicate our earlier findings at 20 and 40 trials and extend them to 120 trials. It is also clear from observing Figure 2 that there is little evidence that the initial relatively high estimates found in Treatment $A(.75 - .25)/B(0 - 1)$ were transient and that ultimately this treatment produces discounting. In support of this observation, none of the interactions involving the three factors was reliable [maximum $F(2,34) = 1.72$].

Plane ratings. The mean effectiveness ratings for the spotter plane contingencies are reported in Table 4. Estimates were clearly in line with the actual contingencies, with little or no fluctuation over trials. Thus, the perfectly negative contingency was rated as strongly negative and the perfectly positive contingency was rated as strongly positive; the mean ratings for the zero contingency hovered around zero. A two-factor repeated measures ANOVA on these data produced a reliable main effect of treatment [$F(2,36) = 364$], but neither the trials main effect nor the interaction was reliable (both $F_s < 1$).

Discussion

Judgments of the camouflage contingency were influenced by the contingency between the second predictor variable (the plane) and the common outcome (safety). In the presence of a perfectly positive plane contingency [$A(.5 - 0)/B(1 - 0)$], the subjects judged the contingency between the camouflage and safety to be lower than in the control treatment [$A(.5 - 0)/B(.25 - .25)$], whereas the presence of a perfectly negative plane contingency caused them to attribute more causal importance to the camouflage than they did in the control treatment. These effects were most certainly not transient and were at least as strong after 120 trials as after 40. The R-W model predicts that the camouflage contingency should be discounted in the presence of a strongly positive plane-

safety contingency, and judgments conformed to this prediction. Although the model predicts a transient increase in estimates, it predicts that in the presence of a perfectly negative plane–safety contingency, the moderate camouflage contingency should be progressively discounted over time. Yet we observed the opposite result.

It might be argued that even with 120 trials the model had not reached asymptote. This might be true if very small learning rate parameters for the context were used or with other parametric manipulations. A strong case can be made against this claim. First, subjects' estimates of intermediate contingencies, which should not have floor or ceiling problems, usually reach an asymptote in fewer than 20 trials (see, e.g., Baker et al., 1989; Shanks, 1985; Shanks, López, Dickinson, & Darby, 1996). Second, the subjects found 120 trials to be very long. Third, if we had chosen parameters that would have delayed asymptote well past 120 trials in order to explain the present result for the $A(.5 - 0)/B(0 - 1)$ treatment, these parameters would have failed to model the perfect estimates of B and many of our other findings, such as the temporal course of learning curves and the course of discounting in other treatments [e.g., that found in the $A(.75 - .25)/B(0 - 1)$ treatment; Baker et al., 1993, and the present Experiment 1]. Moreover, as Figure 1 quite clearly shows, the model predicts the period in which facilitation occurs to be very brief, but the period of enhancement was actually quite extended. Finally, Figure 1 also shows that the model predicts that the judgments from Experiments 1 and 2 should be very similar, including a smaller but parallel enhancement of judgments in Treatment $A(.75 - .25)/(0 - 1)$ relative to its control, $A(.75 - .25)/(.5 - .5)$. Most clearly, this prediction is not confirmed [cf. judgments of $A(.75 - .25)/B(0 - 1)$ and $A(.5 - 0)/B(0 - 1)$ in Figure 2].

Cheng and Holyoak's (1995) PCM also has difficulty with some aspects of our data. This is shown in Table 1. The PCM predicts, as was observed, that judgments in Treatment $A(.5 - 0)/B(.25 - .25)$ should be a function of the camouflage's actual contingency because the overall camouflage contingency and each conditional contingency equals .5. In Treatments $A(.5 - 0)/B(1 - 0)$ and $A(.5 - 0)/B(0 - 1)$, the independent causal status of the camouflage can only be partially assessed because some of the contrasts cannot be computed. Nevertheless, in Treatment $A(.5 - 0)/B(1 - 0)$, although the causal status of the camouflage in the presence of the plane cannot be assessed because there were no plane-present trials without camouflage, the effectiveness of the camouflage in the absence of the plane can be assessed, and it is zero. Melz et al. (1993) have suggested that to assess the causal power of a cause, conditional tests, "to be maximally informative" (p. 1403), must be done in the absence of a known second cause. Cheng et al. (1996) extended this argument by pointing out that excitatory causes may be best understood in the absence of the conditionalizing cue and that inhibitory causes are best understood in the

presence of the conditionalizing excitatory cue. Thus because the conditional contrast for the camouflage in the absence of the plane is defined and equals zero, it can be argued that the PCM predicts discounting in this treatment. The case with the $A(.5 - 0)/B(0 - 1)$ contingency is not so clear. Because $\Delta P_{(\text{camouflage} | \text{no plane})}$ is not defined for this treatment, the subject might be expected to rely on $\Delta P_{(\text{camouflage} | \text{plane})}$, that is, zero, so discounting might be expected. Alternatively, the subject might be confused by the undefined $\Delta P_{(\text{camouflage} | \text{no plane})}$ and might guess. Finally, the subjects might choose a different focal set on which to base their calculation of $\Delta P_{(\text{camouflage} | \text{no plane})}$. Although one of the main explanations of discounting involves the lack of certainty involving undefined contingencies in some discounting preparations (see Waldmann & Holyoak, 1992), associative models do not have this difficulty because they are driven by contiguity, not calculations of covariation. Our subjects may be particularly uncertain with undefined ΔP s, but we have observed no evidence of this uncertainty.

EXPERIMENT 3

A final result of our earlier research that was consistent with a transient but not an asymptotic prediction from the R–W model was our finding that in a treatment where predictor A 's contingency was 0 and predictor B 's contingency was 1 [Treatment $A(.5 - .5)/B(1 - 0)$], judgments of A 's contingency were lower than in a treatment where both predictors had a zero contingency [Treatment $A(.5 - .5)/B(.5 - .5)$; Baker et al., 1993; Vallée-Tourangeau et al., 1994]. According to the model (see Figure 1, right panel), in Treatment $A(.5 - .5)/B(.5 - .5)$, both predictors can develop a small amount of associative strength early on in training (on the basis of a few pairings with the outcome) but will eventually lose it to the context, since the context is a better predictor of the occurrence of the outcome. At asymptote, the associative strength for predictors A and B will equal zero. In Treatment $A(.5 - .5)/B(1 - 0)$, predictor A will also initially acquire some associative strength but will lose this associative strength more rapidly because predictor B rapidly acquires associative strength. Thus the discounting of A in Treatment $A(.5 - .5)/B(1 - 0)$ occurs because A reaches an asymptote of zero more rapidly than in Treatment $A(.5 - .5)/B(.5 - .5)$. Importantly, at asymptote, predictor A should have no associative strength in either Treatment $A(.5 - .5)/B(1 - 0)$ or Treatment $A(.5 - .5)/B(.5 - .5)$. The PCM predicts that predictor A should be judged at zero since its conditionalized ΔP s in the presence and absence of predictor B in both treatments equal zero (Table 1).

Experiment 3 was designed to investigate the long-term stability of discounting found in the $A(.5 - .5)/B(1 - 0)$ treatment. The two critical treatments were $A(.5 - .5)/B(1 - 0)$ and $A(.5 - .5)/B(.5 - .5)$. In addition to these treatments, we included Treatment $A(.75 - .25)/B(.5 -$

.5). This was done to ensure that subjects could discriminate between the zero and a moderately positive contingency for A .

Method

Subjects. Eighteen undergraduates from McGill University volunteered for this experiment.

Design and Procedure. Three treatments were designed by pairing one of two camouflage contingencies ($\Delta P = .5$ or 0) with one of two plane contingencies ($\Delta P = 0$ or 1). In Treatment $A(.75 - .25)/B(.5 - .5)$, the moderately positive camouflage contingency [$p(\text{safety} | \text{camouflage}) = .75, p(\text{safety} | \text{no camouflage}) = .25$] was paired with a zero plane contingency [$p(\text{safety} | \text{plane}) = p(\text{safety} | \text{no plane}) = .5$]. In Treatment $A(.5 - .5)/B(.5 - .5)$, a zero camouflage contingency was paired with the zero plane contingency. And in Treatment $A(.5 - .5)/B(1 - 0)$, the zero camouflage contingency was paired with a perfect plane contingency [$p(\text{safety} | \text{plane}) = 1, p(\text{safety} | \text{no plane}) = 0$]. The nominal contingency tables and conditional contingency tables for this experiment are reported in Table 2.

The same procedure as in Experiment 1 was used. Each treatment lasted 120 trials, and the order of presentation was randomized for each subject. Subjects camouflaged the tank on 61% of the trials. Subjects rated the effectiveness of the camouflage and of the plane after 20, 40, and 120 trials.

Results

Camouflage ratings. Two of the subjects in this experiment did not camouflage the tank on any trial in one or more games. These subjects were eliminated from the analysis reported here. The mean effectiveness ratings for the camouflage are reported in Table 3 and in Figure 2 (left panel). It will be recalled that the R-W model predicted that there should be some initial "discounting" caused by slightly positive judgments in Treatment $A(.5 - .5)/B(.5 - .5)$. These estimates decline until they equal those of Treatment $A(.5 - .5)/B(1 - 0)$. The means shown in Figure 1 are consistent with this prediction, but the supporting statistical evidence is not strong. In Treatment $A(.75 - .25)/B(.5 - .5)$, estimates were at or above 40 throughout the task. In Treatment $A(.5 - .5)/B(.5 - .5)$, ratings were slightly positive after 20 and 40 trials but were slightly negative after 120 trials. In Treatment $A(.5 - .5)/B(1 - 0)$, ratings were slightly negative throughout the task. A two-factor repeated measures ANOVA on these ratings showed that both the main effect of treatment and the interaction were reliable [$F(2,30) = 16.1, F(4,60) = 2.88$], but the main effect of trials ($F < 1$) was not.

In the analysis comparing estimates of the control treatment $A(.75 - .25)/B(.5 - .5)$ and the $A(.5 - .5)/B(.5 - .5)$ treatment, the main effect and the interaction were reliable [$F(1,15) = 21.0$, and $F(2,30) = 4.61$], and the effect of trials was not [$F(2,30) = 1.13$]. This analysis confirms that the subjects discriminated the moderately positive and zero camouflage contingencies throughout; this is consistent with the apparent decline in estimates of Treatment $A(.5 - .5)/B(.5 - .5)$. Estimates of Treatments $A(.75 - .25)/B(.5 - .5)$ and $A(.5 - .5)/B(.5 - .5)$ did not differ reliably after 20 trials [$F(1,15) = 3.76$], but did on the last two blocks of trials [$F(1,15) = 21.6$]. The analysis comparing the $A(.5 - .5)/B(.5 - .5)$ treatment and the

$A(.5 - .5)/B(1 - 0)$ treatment revealed no reliable effects, although the difference between treatments was nearly reliable [$F(1,15) = 3.95, p < .07$]. The interaction was not reliable, and this null result is not consistent with the prediction and the trend that the means of these two treatments should approach one another [$F(2,30) = 2.12, p < .15$]. Because trial-by-trial comparisons of these two treatments were preplanned, rejecting the R-W model requires accepting the null, and because there was a reliable interaction in the overall ANOVA, we carried out the more powerful trial-by-trial comparison of these two treatments. The two treatments differed reliably over the first two blocks of trials [$F(1,15) = 5.97$], but not after 120 trials [$F(1,15) < 1$]. The second set of analyses, justified by the significant interaction in the overall ANOVA, involved three within-treatments comparisons over the three estimation periods. These analyses provided very weak evidence for a decline in Treatment $A(.5 - .5)/B(.5 - .5)$, since the appropriate F ratio was not quite reliable [$F(2,30) = 2.83, p > .08$].

Plane ratings. The mean effectiveness ratings for the plane are reported in Table 4. In both Treatments $A(.75 - .25)/B(.5 - .5)$ and $A(.5 - .5)/B(.5 - .5)$, estimates hovered around zero, whereas in Treatment $A(.5 - .5)/B(1 - 0)$, the mean terminal estimate was 97.1 [$F(2,30) = 27.8$].

Discussion

The results and analysis of this experiment provide weak support for the predictions of the R-W model. As the model predicts, the subjects do discriminate between zero camouflage contingencies and moderately positive ones. Furthermore, there was an interaction between treatments over time, and there was some statistical evidence that this interaction occurred because initial estimates of the $A(.5 - .5)/B(.5 - .5)$ contingency were positive and these estimates declined. Because this comparison was preplanned, we used powerful and often redundant comparisons that were sometimes not justified by reliable interactions. In spite of this, an equally strong statistical case could be made for the claim that there was little decline in discounting over the 120 trials. Moreover, this decrease in estimates occurred over many trials, and the parameters necessary to predict this interaction would have difficulty predicting our other effects (see, e.g., Baker et al., 1993).

As noted, the PCM predicts no discounting in this experiment. However, if judgments were based on a different focal set including time between trials, then some of the results here might be expected (see Vallée-Tourangeau, Murphy, & Baker, 1996).

GENERAL DISCUSSION

We have reported three experiments designed to investigate subjects' estimates of contingency once asymptote was achieved in a relative validity preparation (Baker et al., 1993; Wagner et al., 1968). In Experiment 1, we

found that judgments of a moderate-density, moderately positive camouflage were discounted when the camouflage was paired with a perfectly positive or perfectly negative plane contingency. In this experiment, discounting was stronger with the perfectly positive plane. Furthermore, discounting seemed to develop over trials. In Experiment 2, we found that the higher estimates of the low-density, moderately positive contingency when it was paired with a perfectly negative plane [$A(.5 - 0)/B(0 - 1)$], relative to the control treatment [$A(.5 - 0)/B(.25 - .25)$], were not transient. In the third experiment, there was much weaker evidence that the reduction in judgments found with a zero camouflage contingency in the presence of a perfect plane contingency was transient.

When we initially published similar results using only 40 trials, we were very sanguine about the R-W model as an explanation of the judgment process (Baker et al., 1993). Certain other aspects of the data also seemed to fit the model well. These included unpublished attempts to get our student subjects to articulate the cognitive processes that they used to make their judgments. They generally could not articulate a process but would argue that they were simply guessing or that they just "felt" their judgment was appropriate. More objective data came from experiments asking for judgments in a situation using geometric symbols and no causal scenario (see Vallée-Tourangeau et al., 1994). Furthermore, Wasserman et al. (1993) reported that the R-W model provided a very good model of judgments in an operant contingency task. Moreover, the subjective judgments of the conditional probabilities—namely $p(\text{outcome} | \text{response})$ and $p(\text{outcome} | \text{no response})$ —did not predict their judgments of ΔP as well as did the R-W model (Wasserman et al., 1993). In addition, Shanks (1989) has provided data that have shown the importance of contiguity in judgments of causality. Nonetheless, it should be emphasized that each of the preceding arguments applies to the associative approach in general and not specifically to the R-W model.

Yet the data from the present experiments are at best mixed in their support of the R-W model as a unitary explanation. In Experiment 1, although discounting should have taken longer to develop in the negative discounting treatment [$A(.75 - .25)/B(0 - 1)$] than in the positive discounting treatment [$A(.75 - .25)/B(1 - 0)$], the model predicts that discounting will ultimately be equal at asymptote. This latter prediction was not confirmed in this experiment. In Experiment 2, the prolonged period in which judgments of the camouflage in the low-density negative discounting treatment [$A(.5 - 0)/B(0 - 1)$] were elevated above the control treatment [$A(.5 - 0)/B(.25 - .25)$] was not expected. In Experiment 3, the model's prediction that the "discounting" found in the $A(.5 - .5)/B(1 - 0)$ treatment dissipates because the initial excitation in the $A(.5 - .5)/B(.5 - .5)$ treatment takes longer to extinguish, received only equivocal support.

However, an alternative associative model, Pearce's (1987) stimulus generalization model, offers a closer fit of these data. Like the R-W model, the Pearce model can be thought of as a simple adaptive network, but the input

layer represents configurations of stimuli experienced by the subjects, not the individual candidate causes. In the experiments reported here, subjects experienced four configural stimuli, namely "camouflage + plane + context" (ABX), "camouflage + context" (AX), "plane + context" (BX), and "context" (X). Each configuration is assigned a unit in the input layer, and each of those units is linked to the output layer (representing the outcome) via an excitatory connection and an inhibitory connection. Changes in the excitatory (E) strength of each configuration is a function of

$$\Delta V = \beta(\lambda - \bar{V}),$$

where \bar{V} refers to the net associative strength or the sum of the excitatory and inhibitory strength of the configuration, which in turn includes generalized excitatory and inhibitory potential from similar configurations. For example, the net associative strength of ABX includes not only its own excitatory and inhibitory strength, but also generalized excitatory and inhibitory strength, from AX , BX , and X . The amount of associative strength generalized is a function of the similarity of that configuration to ABX ; for example, the coefficient that indexes the similarity of AX to ABX is .67 (see Pearce, 1987, Equation 5). Cue competition is thus realized in the Pearce model indirectly, via the generalized strength of similar cues, not by the sum of the associative strength of cues present on a trial. Since A always occurs either within the configuration ABX (camouflage + plane + context) or AX (camouflage + context), A 's predicted effectiveness rating is a function of the sum of the excitatory (E) and inhibitory (I) connections of ABX and AX weighted by similarity coefficients, or

$$[.33(E_{ABX} + I_{ABX}) + .5(E_{AX} + I_{AX})].$$

The mean predicted strengths of A for each treatment in each of the three experiments are plotted in Figure 3. The predictions shown in the left panel reflect the two phenomena observed in Experiment 1, namely faster blocking in $A(.75 - .25)/B(1 - 0)$ (triangles) than in $A(.75 - .25)/B(0 - 1)$ (filled circles) and, after 120 trials, the asymptotic blocking is less in $A(.75 - .25)/B(0 - 1)$ than in $A(.75 - .25)/B(1 - 0)$. The middle panel (Experiment 2) shows that the Pearce model does not predict blocking in $A(.5 - 0)/B(0 - 1)$ (filled circles) and, indeed, predicts higher ratings than in $A(.5 - 0)/B(.25 - .25)$. This is what was observed. Blocking is predicted in $A(.5 - 0)/B(1 - 0)$. As for Experiment 3 (right panel), faster blocking of the camouflage in $A(.5 - .5)/B(.5 - .5)$ than in $A(.5 - .5)/B(1 - 0)$ is predicted by the Pearce model, although the ratings do not converge with training. Overall, the fit between the predictions derived from the Pearce model and the ratings is impressive. Recent efforts suggest that the Pearce model accounts for other aspects of human causal induction as well (e.g., López, Shanks, Almaraz, & Fernández, 1998; Vallée-Tourangeau, Murphy, Drew, & Baker, 1998).

The conclusion with the most apparent parsimony that can be drawn from our results is that neither the R-W

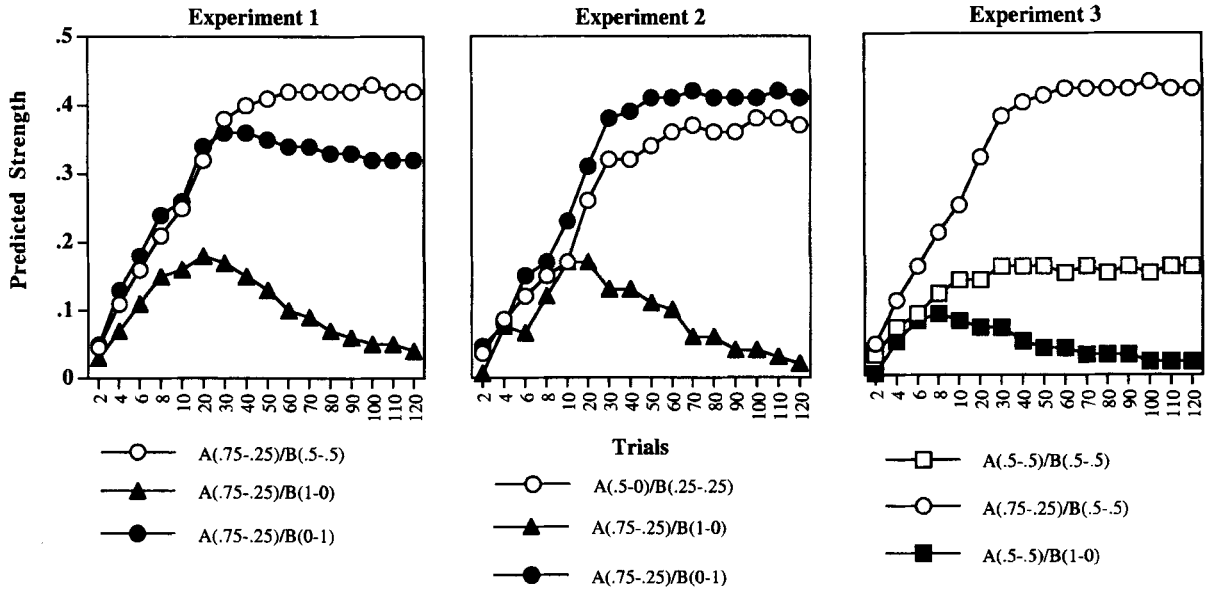


Figure 3. Mean predicted effectiveness ratings for the camouflage derived from the Pearce (1987) model in each of the three treatments of Experiments 1, 2, and 3. $\beta = .25$ for all trials; means calculated over 30 iterations.

model nor the PCM accounts well for the dynamic properties of discounting or blocking found with our preparation. Furthermore, it would seem that Pearce's associative model provides a better fit of our data. In the introduction we mentioned that, beyond the question of which specific model provided a better account of causal reasoning, there was a more general friction between the general classes of theories we called associative and the statistical or normative theories. Our results would seem to support the former class of model, yet we feel this conclusion is premature for three reasons. First, as we have argued elsewhere (Baker, Murphy, & Vallée-Tourangeau, 1996), causal reasoning is multifaceted and the brain is certainly modular, so there is room for both types of models to co-exist. Second, it is impossible to rule out any general class of model with a finite data set. The PCM is a specific statistical model, and although it does not model our data well, it is, in principle, possible to generate a statistical model that accounts for our results. Post hoc selection of models is not particularly parsimonious, but it must be remembered that that is what we did within the associative framework when we switched from the R-W model to Pearce's model.

Third, and most critical, cognitive or statistical models not only represent possible psychological mechanisms for causal reasoning but they also represent a rational analysis of the environmental contingencies. The output of these rational calculations represents a series of ideals that the organism or system must compute if its reasoning is to be adaptive. Because many phenomena such as blocking or discounting are the consequence of these calculations (e.g., discounting in the $A(.75 - .25)/B(1 - 0)$ contingency of Experiment 1 is described by

conditional ΔPs), these phenomena may themselves be normative and not examples of failures to learn or other errors of the reasoner. This supports the argument that the cognitive processes studied within our somewhat arbitrary preparations are, indeed, adaptive and not intellectual curiosities or epiphenomena. Thus, at the very least, the normative models represent an ideal and possibly a computational model of the mind.

At a deeper, more mechanistic level, it is possible that models like the PCM describe an irreducible algorithm for causal reasoning. However, it is also possible that associative mechanisms are used. From our reductionist position, we are biased in favor of the latter. We believe that ultimately behavior must be modeled by mechanisms that do not include direct normative representations of events and calculations based on these representations. After all, brains are constructed of neurons.

At the very least, it seems reasonable to assume that both associative and normative mechanisms complement each other at different levels of analysis. As scientists we need both. Moreover, both our cognitive world and our brains are modular. It is likely that some module might be more amenable to one analysis or the other. Each class of model might provide a mechanism for different cognitive calculations. Within the contingency learning preparations we used here, associative mechanisms seem more appropriate.

REFERENCES

ALLAN, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, 15, 147-149.
 BAKER, A. G., BERBRIER, M., & VALLÉE-TOURANGEAU, F. (1989). Judg-

- ments of a 2×2 contingency table: Sequential processing and the learning curve. *Quarterly Journal of Experimental Psychology*, **41B**, 65-97.
- BAKER, A. G., MERCIER, P., VALLÉE-TOURANGEAU, F., FRANK, R., & PAN, M. (1993). Selective association and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 414-432.
- BAKER, A. G., MURPHY, R. A., & VALLÉE-TOURANGEAU, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D. R. Shanks, K. Holyoak, & D. Medin (Eds.), *The psychology of learning and motivation* (Vol. 34, pp. 1-45). San Diego: Academic Press.
- CHAPMAN, G. B., & ROBBINS, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, **18**, 537-545.
- CHENG, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, **104**, 367-405.
- CHENG, P. W., & HOLYOAK, K. J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In J. A. Meyer & H. Roitblat (Eds.), *Comparative approaches to cognitive science* (pp. 271-302). Cambridge, MA: MIT Press.
- CHENG, P. W., & NOVICK, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, **99**, 365-382.
- CHENG, P. W., PARK, J., YARLAS, A. S., & HOLYOAK, K. J. (1996). A causal-power theory of focal sets. In D. R. Shanks, K. Holyoak, & D. Medin (Eds.), *The psychology of learning and motivation* (Vol. 34, pp. 313-355). San Diego: Academic Press.
- 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.
- KAMIN, L. J. (1969). Selective association and conditioning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 42-64). Halifax, NS: Dalhousie University Press.
- LÓPEZ, F. J., SHANKS, D. R., ALMARAZ, J., & FERNÁNDEZ, P. (1998). Effects of trial order on contingency judgments: A comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 672-694.
- MELZ, E. R., CHENG, P. W., HOLYOAK, K. J., & WALDMANN, M. R. (1993). Cue competition in human categorization: Contingency or the Rescorla-Wagner learning rule? Comment on Shanks (1991). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1398-1410.
- PEARCE, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, **94**, 61-73.
- PRICE, P. C., & YATES, F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition*, **21**, 561-572.
- PRICE, P. C., & YATES, F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1639-1655.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- SHANKS, D. R. (1985). Continuous monitoring of human contingency judgment across trials. *Memory & Cognition*, **13**, 158-167.
- SHANKS, D. R. (1989). Selectional processes in causality judgment. *Memory & Cognition*, **17**, 27-34.
- SHANKS, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 433-443.
- SHANKS, D. R., LÓPEZ, F. J., DICKINSON, A., & DARBY, R. J. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, K. Holyoak, & D. Medin (Eds.), *The psychology of learning and motivation* (Vol. 34, pp. 229-261). San Diego: Academic Press.
- VALLÉE-TOURANGEAU, F., BAKER, A. G., & MERCIER, P. (1994). Discounting in causality and covariation judgments. *Quarterly Journal of Experimental Psychology*, **47B**, 151-171.
- VALLÉE-TOURANGEAU, F., MURPHY, R. A., & BAKER, A. G. (1996). Judging the contingency of a constant cause: Contrasting predictions from an associative and a statistical model. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 447-452). Mahwah, NJ: Erlbaum.
- VALLÉE-TOURANGEAU, F., MURPHY, R. A., DREW, S., & BAKER, A. G. (1998). Judgments of constant and variable candidate causes: A test of the PowerPC theory. *Quarterly Journal of Experimental Psychology*, **51A**, 65-84.
- WAGNER, A. R., LOGAN, F. A., HABERLANDT, K., & PRICE, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, **76**, 171-180.
- 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., ELEK, S. M., CHATLOSH, D. J., & BAKER, A. G. (1993). Rating causal relations: The role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 174-188.

(Manuscript received November 17, 1997;
revision accepted for publication December 27, 1998.)