Available online at www.sciencedirect.com
ScienceDirect
Behavioural Processes 74 (2007) 251–264
BEHAVIOURAL PROCESSES
www.elsevier.com/locate/behavproc

Judgments of causal efficacy under constant and changing interevent contingencies

Masayoshi Katagiri[a,1], Shu-Fang Kao[b], Adam M. Simon[c], Leyre Castro[d], Edward A. Wasserman[d,*]

[a] Utsunomiya University, Utsunomiya, Japan
[b] National Hsinchu University of Education, Hsinchu City, Taiwan, ROC
[c] Fiedler Center at Northwestern University, Evanston, IL, United States
[d] Department of Psychology, E11 Seashore Hall, The University of Iowa, Iowa City, IA 52242-1407, United States

Received 9 June 2006; received in revised form 6 September 2006; accepted 6 September 2006

Abstract

How do people judge constant and varying interevent contingencies? In two experiments, 150 college students rated the efficacy of a potential cause (an experimental fertilizer) of an effect (a plant's blooming). The prevailing probabilistic interevent relation could remain constant for the entirety of the problem or it could change without warning at the midway point: by contingency reversal, by shifting from noncontingency to contingency, or by shifting from contingency to noncontingency. Participants' trial-by-trial ratings sensitively tracked the prevailing positive, negative, and noncontingent interevent relations, even those that entailed an unsignaled change in contingency. Changes in specific cells of the 2×2 contingency table differentially affected participants' response to the altered interevent relations. All of this evidence was well described by an associative account of contingency and causal judgments.

Keywords: Contingency judgments; Causal learning

One of the most daunting cognitive challenges facing human and nonhuman animals alike is to detect and to respond effectively to those environmental events that are truly contingent on other events or behaviors. Sensing interevent relations when they are absent or failing to sense interevent relations when they are present could be catastrophic to organisms' responding adaptively to their intricate and constantly changing environments.

Early in his illustrious scholarly career, Russell Church was especially concerned with separating the effects of contingent punishment from noncontingent punishment. Church was to summarize the results of his and others' research with the observation that: "The magnitude of suppression of an instrumental response is greater if the punishment is contingent upon the response than if it is not" (Church, 1969, p. 132). In conjunction with this work, Church (1964) also inquired into the logical status of the yoked control procedure, a method frequently used to equalize the number and temporal distribution of shocks received by organisms under contingent and noncontingent conditions (for more on Church's analysis, see Wasserman, 1988).

These interests have since receded in importance to Church's own research program, but they have moved to the forefront in contemporary psychological science. A lively realm of inquiry now revolves around organisms' discriminating contingency from noncontingency, where the focal relations are between different environmental events or between environmental events and the behaviors that may have produced them or prevented their occurrence.

Generally, experimental studies of interevent relations entail the factorial combination of the occurrence/nonoccurrence of a stimulus or a behavior with the occurrence/nonoccurrence of a certain effect, which yields four possible types of interevent information: Cell A, both events present; Cell B, stimulus or behavior present and effect absent; Cell C, stimulus or behavior absent and effect present; and Cell D, both events absent. There are a variety of measures to calculate the contingency between events; in predictive tasks, where we are interested in whether or not the effect is dependent on the occurrence of the stimulus or behavior, contingency is normally calculated using the Δp rule

* Corresponding author. Tel.: +1 319 335 2445.
E-mail address: ed-wasserman@uiowa.edu (E.A. Wasserman).

0376-6357/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.beproc.2006.09.001

(see Allan, 1980):

$$\Delta p = \frac{A}{A+B} - \frac{C}{C+D}$$

Consider some illustrations of contingency discrimination in our daily lives. Most of us have had the discomforting experience of early empirical evidence suggesting a relation between events that eventually turned out not to hold true in the long run. An innovative sales campaign might at first seem to be more profitable than the prior one; submitting papers to a new journal might initially appear to be more successful than submitting them to an established periodical; or testing an experimental drug might at first seem to be more therapeutic than a well-ensconced drug treatment. There seems to be a tendency to give more relevance to initial information—a *primacy* effect. It is precisely because of our possible predisposition to form premature impressions of that evidence that we are admonished by statisticians to wait until a large sample of data has been collected and a proper statistical analysis has been conducted before deciding whether or not a reliable relation holds between the variables of interest.

Behavioral evidence consistent with such a primacy effect was first reported by Yates and Curley (1986) in the contingency scores that they *derived* from their participants' conditional probability ratings. More recently, Dennis and Ahn (2001) also reported a primacy effect when participants were required to make *direct* contingency ratings. However, Kao (1993) and Wasserman et al. (1996) had earlier found no hint of primacy in their research; their participants' contingency ratings closely accorded with the overall interevent contingencies that they were given. To further complicate the matter, López et al. (1998) reported reliable *recency* effects. All of these studies used interevent relations in which the first half of the evidence suggested one relation, but the second half of the evidence suggested exactly the opposite relation. So, overall, there was no statistical association between the binary variables.

Moving to a different, yet related experimental design, early evidence does exert a clear and disproportionate effect on the psychological impact of *multiple* cues in contingency judgment and associative learning studies. For example, Stimulus A blocks the control acquired by Stimulus X when Stimulus A alone is paired with an outcome prior to pairing the AX compound with that outcome. This so-called "forward blocking" effect is measured against the behavioral control acquired by Stimulus X when only the AX compound is paired the same number of times with the outcome. Forward blocking has been found for humans rating interevent relations (e.g., Chapman, 1991; Chapman and Robbins, 1990; Dickinson et al., 1984; Shanks, 1985) or performing behavioral tasks (e.g., Arcediano et al., 1997) as well as for rats learning CS-US contingencies in Pavlovian conditioning situations (e.g., Kamin, 1969; Miller and Matute, 1996).

One prominent interpretation of the forward blocking effect follows from the influential model of associative learning proposed by Rescorla and Wagner (1972). The model argues that prior training with Stimulus A effectively consumes all of the associative strength that the outcome has available before the AX compound is ever given. When it is given, Stimulus X fails to form the associative connection with the outcome that it would have formed had pairings of Stimulus A with the outcome not preceded pairings of AX with the outcome.

Although the Rescorla–Wagner model does predict a deleterious effect of prior training on novel added stimuli—as in the case of blocking, the model does *not* predict that primacy will result from the reversed-contingency training procedures in which a certain cue-outcome contingency is suggested in a first phase and the opposite contingency is suggested in a second phase (Dennis and Ahn, 2001; Kao, 1993; López et al., 1998; Yates and Curley, 1986; Wasserman et al., 1996). Depending on the parameters that are input into the model, the Rescorla–Wagner model either predicts that primacy will not occur or that recency will occur (Kao, 1993). The prediction of recency rather than primacy derives from the fact that *changes* in associative strength due to contingency reversal are of greater magnitude the farther the present level of associative strength is from zero. Predisposing a nonzero relation during the first half of training – as in the studies of Yates and Curley (1986) and Kao (1993) – means that reversing the interevent contingency will either equivalently reverse the level of associative strength or that reversal will lead to overweighting of the associative relation implied by the *latter* half of the interevent information.

The present pair of experiments further pursued the possibility of primacy (Experiment 1) and the possible differential impact of each of the four 2 × 2 contingency cells in a causal judgment task (Experiment 2). We systematically manipulated interevent relations both between and within individual problems that participants were given. In both investigations, many different contingency judgment problems were presented to participants, in which the interevent relation holding during the first half of the problem either continued to hold or changed without warning during the second half of the problem. In Experiment 1, the within-problem changes in contingency: (a) reversed the relation prevailing during the first half, (b) shifted from a contingent to a noncontingent relation, or (c) shifted from a noncontingent to a contingent relation. In Experiment 2, the within-problem changes in contingency: (a) reversed the relation prevailing during the first half or (b) shifted from a contingent to a noncontingent relation. In Experiment 1, the changes in contingency were effected by equivalent increases and decreases in the contents of all four cells of the 2 × 2 contingency table. The changes in contingency in Experiment 2 were effected by changes in the frequencies of either two or four cells of the 2 × 2 contingency table. Based on a large body of evidence indicating that the psychological impact of the four evidential cells is ordered Cell A > Cell B > Cell C > Cell D (Catena et al., 1998; Kao and Wasserman, 1993; Levin et al., 1993; Mandel and Lehman, 1998; Schustack and Sternberg, 1981; Wasserman et al., 1990; White, 2003), we made detailed predictions about just which changes in cell frequencies would have the greatest effect on participants' ratings during the second half of the problems with changed contingencies. Those differential cell weightings had to be considered in light of the fact that entries in Cells A and D are consistent with positive interevent relations, whereas entries in Cells B and C are consistent with negative interevent relations.

The results of these two experiments combined not only attest to the remarkable ability of college students to track the interevent contingencies that they are given, but they also underscore the utility of associative learning accounts, such as the Rescorla–Wagner model, to accommodate the momentary modulations in contingency judgments that participants exhibit.

1. Experiment 1

The participants in our first experiment were given a total of 13 different contingency problems to rate: 5 involved interevent relations that remained the same throughout the 24 trials of each problem and 8 involved interevent relations that changed without warning at the midway point. This set of contingency problems afforded us the opportunity to examine participants' general sensitivity to the relations that they were given as well as the chance to see whether participants' judgments were disproportionately influenced by the early interevent information to which they were exposed—primacy.

In other judgment tasks, it has been found that participants' ratings may differ depending on whether they make ratings every time a new piece of information is presented or whether they make only a single rating at the end of the iterated information (Arkes and Harkness, 1983; Catena et al., 1998; Collins and Shanks, 2002; Hastie and Park, 1986; Matute et al., 2002). Specifically, some studies have reported that frequent judgments during the training phase induce recency effects, whereas only one judgment at the end of the training phase induces a more accurate judgment of the overall contingency (Catena et al., 1998; Collins and Shanks, 2002; Matute et al., 2002). However, participants in these studies were not given instructions as to the information that they should consider when making their ratings: either all of the information presented to that point or just the latest piece of information. It is conceivable that the reported recency effects were due merely to the participants' confusion as to the information that should be rated; focusing on just the last piece of information would artifactually suggest a recency effect. In order to control for this possibility, we had two different groups of participants make continuous or terminal contingency judgments and we included specific instructions in which participants were told to take into account all of the information that had been presented to that point, not just the last piece of information (see instructions below).

2. Method

2.1. Participants

The participants were 100 (50 male and 50 female) students in introductory psychology classes at the University of Iowa, who received course credit for their participation.

2.2. Apparatus

The participants were tested individually with one of two Hewlett Packard Vectra 486/33U computers in the same large room, each computer equipped with a color CRT monitor. Programs were developed in Pascal.

2.3. Design

The experiment involved 13 different contingency problems that are shown in Table 1. These 13 problems can be organized into five different groupings in order to clarify the behavioral effects of those problems that either did or did not involve an unsignaled change in contingency at the halfway point. All of the 24-trial problems were specified by the contents of the 2×2 information matrix (Cell A, Cell B, Cell C, and Cell D) scheduled during the first and second halves of the problems. Individual cell entries were either 1, 2, 3, 4, or 5, with the constraint that the sum of the four cells was 12 in each half of each problem. These cell frequencies yielded Δps whose absolute values were .000, .333, or .667 in either half of the problems or over both halves of the problems. The five groupings were:

1. *Problems with the same contingency in each half of the 24 trials*. This group included Problems 1, 2, 3, 4, and 5, entailing Δps of +.667, +.333, .000, −.333, and −.667, respectively.
2. *Problems with an overall Δp of zero*. This group included: Problem 3 with $\Delta p = .000$ in both halves; Problem 6 with $\Delta p = +.667$ in the first half and $\Delta p = -.667$ in the second; Problem 9 with $\Delta p = -.667$ in the first half and $\Delta p = +.667$ in the second; Problem 7 with $\Delta p = +.333$ in the first half and $\Delta p = -.333$ in the second; and Problem 8 with $\Delta p = -.333$ in the first half and $\Delta p = +.333$ in the second.
3. *Problems with an overall absolute value of $\Delta p = |.333|$*. Problems 2, 10, and 11 had an overall $\Delta p = +.333$; Problems 4, 12, and 13 had an overall $\Delta p = -.333$. Problems 2 and 4 entailed no change in contingency; Problems 10 and 12 entailed a change from a nonzero to a zero contingency; and Problems 11 and 13 entailed a change from a zero to a nonzero contingency.
4. *Problems with a zero contingency in the second half* (Problems 3, 10, and 12).
5. *Problems with a zero contingency in the first half* (Problems 3, 11, and 13).

2.4. Experimental conditions

The between-subject variable manipulated in this study was running versus final contingency estimation. Half of the participants were assigned to the running-estimate condition, in which they had to update their judgments on every trial; the other half of the participants were assigned to the final-estimate condition, in which they made their judgments only before the first trial and after the last trial of each problem. In order to make the independent variables more alike between conditions, each participant in the running-estimate condition was paired with one of the same sex in the final-estimate condition; the same information presentation sequence was thus given to each pair of participants. As well, in order to approximately equalize the duration of the experiment for both conditions, participants in the final-estimate condition had 2 s added to the time between

Table 1
Problems used in Experiment 1 and participants' ratings

Problems	Cell entries												Contingencies			Participants' ratings					
	Block 1				Block 2				Overall				Block 1	Block 2	Overall	Final-estimate condition			Running-estimate condition		
	A	B	C	D	A	B	C	D	A	B	C	D				T0	T24	T24 − T0	T0	T24	T24 − T0
Problems with all contingencies constant																					
1	5	1	1	5	5	1	1	5	10	2	2	10	0.667	0.667	0.667	1.14	6.36	5.22	0.08	5.32	5.24
2	4	2	2	4	4	2	2	4	8	4	4	8	0.333	0.333	0.333	0.82	2.90	2.08	0.30	4.06	3.76
3	3	3	3	3	3	3	3	3	6	6	6	6	0.000	0.000	0.000	1.54	0.24	−1.30	0.66	0.38	−0.28
4	2	4	4	2	2	4	4	2	4	8	8	4	−0.333	−0.333	−0.333	1.22	−2.30	−3.52	0.50	−2.34	−2.84
5	1	5	5	1	1	5	5	1	2	10	10	2	−0.667	−0.667	−0.667	0.96	−5.02	−5.98	0.74	−4.96	−5.70
Problems with zero overall contingency																					
6	5	1	1	5	1	5	5	1	6	6	6	6	0.667	−0.667	0.000	0.66	0.42	−0.24	0.98	−0.42	−1.40
7	4	2	2	4	2	4	4	2	6	6	6	6	0.333	−0.333	0.000	1.22	0.54	−0.68	0.58	−0.12	−0.70
(3)	3	3	3	3	3	3	3	3	6	6	6	6	0.000	0.000	0.000	1.54	0.24	−1.30	0.66	0.38	−0.28
8	2	4	4	2	4	2	2	4	6	6	6	6	−0.333	0.333	0.000	0.22	0.42	0.20	0.52	0.06	−0.46
9	1	5	5	1	5	1	1	5	6	6	6	6	−0.667	0.667	0.000	0.30	0.38	0.08	0.24	1.24	1.00
Problems with contingency of |.333|																					
10	5	1	1	5	3	3	3	3	8	4	4	8	0.667	0.000	0.333	0.16	4.50	4.34	1.36	2.98	1.62
(2)	4	2	2	4	4	2	2	4	8	4	4	8	0.333	0.333	0.333	0.82	2.90	2.08	0.30	4.06	3.76
11	3	3	3	3	5	1	1	5	8	4	4	8	0.000	0.667	0.333	1.04	3.88	2.84	0.42	3.14	2.72
12	1	5	5	1	3	3	3	3	4	8	8	4	−0.667	0.000	−0.333	0.96	−2.78	−3.74	0.82	−2.26	−3.08
(4)	2	4	4	2	2	4	4	2	4	8	8	4	−0.333	−0.333	−0.333	1.22	−2.30	−3.52	0.50	−2.34	−2.84
13	3	3	3	3	1	5	5	1	4	8	8	4	0.000	−0.667	−0.333	0.60	−2.42	−3.02	0.54	−3.02	−3.56
Problems ending with no contingency																					
(10)	5	1	1	5	3	3	3	3	8	4	4	8	0.667	0.000	0.333	0.16	4.50	4.34	1.36	2.98	1.62
(3)	3	3	3	3	3	3	3	3	6	6	6	6	0.000	0.000	0.000	1.54	0.24	−1.30	0.66	0.38	−0.28
(12)	1	5	5	1	3	3	3	3	4	8	8	4	−0.667	0.000	−0.333	0.96	−2.78	−3.74	0.82	−2.26	−3.08
Problems starting with no contingency																					
(11)	3	3	3	3	5	1	1	5	8	4	4	8	0.000	0.667	0.333	1.04	3.88	2.84	0.42	3.14	2.72
(3)	3	3	3	3	3	3	3	3	6	6	6	6	0.000	0.000	0.000	1.54	0.24	−1.30	0.66	0.38	−0.28
(13)	3	3	3	3	1	5	5	1	4	8	8	4	0.000	−0.667	−0.333	0.60	−2.42	−3.02	0.54	−3.02	−3.56

Note: values in parentheses indicate that the problem has already appeared above.

presentations of successive contingency information. Prior work in our laboratory had shown that this time was roughly equal to that taken by running-estimate participants to make their contingency ratings.

Problem was a within-subject variable, so that all of the participants were presented with the 13 problems in a random order; for all 13 problems, the presentation of cell information within each half was random.

2.5. Procedure

After arriving at the laboratory and giving their informed consent, participants were asked to attend to the information on the computer screen and to follow the instructions step-by-step; the instructions for both conditions were as follows:

> Suppose you are employed by the American Flowering Plants Laboratory that has developed 13 experimental fertilizers, which are labeled F1 through F13, for promoting the Lanyu to bloom. The Lanyu, an exotic plant, is imported from Brazil. *It is possible that an experimental fertilizer may promote or suppress the Lanyu's blooming, or it may have no effect at all on the Lanyu's blooming*. Before being introduced to the market, each fertilizer has to be tested and evaluated for its effectiveness. To do so, each fertilizer was tested with a different group of plants. Within each group, the fertilizer was given once to some plants, but not to others. When the fertilizer was given, it was dissolved in a quantity of water; when the fertilizer was not given, the plant was provided with the same quantity of water.

After 2 days, within each test group, four types of plants can be identified:

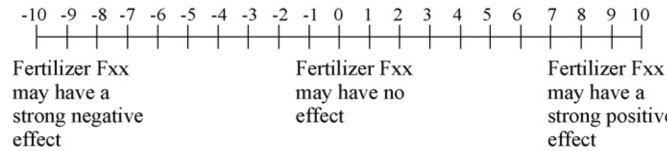
(A) Plant received the fertilizer and bloomed.
(B) Plant received the fertilizer, but did not bloom.
(C) Plant did not receive the fertilizer, but bloomed.
(D) Plant did not receive the fertilizer and did not bloom.

Now, given the test result of each plant in a group, your job is to assess the effectiveness of each tested fertilizer on the Lanyu's blooming.

The evaluation procedure for each fertilizer is as follows:

1. Before being given any test results, you will be asked to give your expectation of the ability of the fertilizer to affect the Lanyu's blooming, rating its possible effectiveness on a scale of −10 to +10:
   A score of −10 means that you think the fertilizer may have a strong negative effect on the Lanyu's blooming.
   A score of 0 means that you think the fertilizer may have no effect on the Lanyu's blooming.
   A score of +10 means that you think the fertilizer may have a strong positive effect on the Lanyu's blooming.
   And, scores between −10 and +10 on the scale indicate different degrees of the fertilizer's effectiveness.

   The response scale will look like:



2. You will then be given a series of descriptions each indicating the state of a Lanyu plant: it either received or did not receive the fertilizer and it either bloomed or did not bloom. The left side of the screen indicates the status of a Lanyu plant before the test, whereas the right side of the screen indicates the status of the same Lanyu plant after the test.
   For example:
   If the plant received the fertilizer and bloomed, then you will see the left side of the screen display "Did the plant receive Fertilizer xx? YES," and the right side display "Did the plant bloom? YES."
   [press SPACEBAR to see what happens . . .]
   If the plant received the fertilizer, but did not bloom, then you will see the left side of the screen display "Did the plant receive Fertilizer xx? YES," and the right side display "Did the plant bloom? NO."
   [press SPACEBAR to see what happens . . .]
   If the plant did not receive the fertilizer, but bloomed, then you will see the left side of the screen display "Did the plant receive Fertilizer xx? NO," and the right side display "Did the plant bloom? YES."
   [press SPACEBAR to see what happens . . .]
   If the plant did not receive the fertilizer and did not bloom, then you will see the left side of the screen display "Did the plant receive Fertilizer xx? NO," and the right side display "Did the plant bloom? NO."
   [press SPACEBAR to see what happens . . .]
   The following instructions were for the running-estimate condition:
3. Before being given a particular description, you will hear a short tone to forewarn you to pay attention to the upcoming information.
4. After seeing a particular description, you will then update your evaluation of the effectiveness of that particular fertilizer. This means that your rating is to reflect the overall influence of all of the plants you have observed to that point—not just the last plant.
5. Note that the position of the cursor along the response scale on each trial, except on the initial guess trial, indicates your judgment from the previous trial so that you can adjust your judgment from there (a short description of the numerical value of your earlier judgment will also appear above the scale). In addition, each time after you make a decision you can modify your decision by pressing the DEL key or confirm your decision by pressing SPACEBAR.
6. Steps 3–5 will be repeated until all test results of a group of plants are given.
   The following instructions were for the final-estimate condition:

3. Before being given a particular description, you will hear a short tone to forewarn you to pay attention to the upcoming information. The presentation interval between pieces of information is 2 s. This step will be repeated until all test results of a group of plants are given.
4. You will then be asked to give your overall evaluation about the effectiveness of that particular tested fertilizer.
5. Note that the position of the cursor along the response scale on the last trial indicates your initial estimate so that you can adjust your judgment from there (a short description of the numerical value of your initial estimate will also appear above the scale). In addition, each time after you make a decision you can modify your decision by pressing the DEL key or confirm your decision by pressing SPACEBAR.

When making their contingency ratings, participants were instructed to use the left and right arrow keys to move the cursor along the response scale. Participants were also given a practice problem with 24 trials to help familiarize them with the procedure; each type of cell information had an equal probability of being presented on each trial. If participants did not have any questions after the practice problem, then the 13 designated contingency problems were given; the session lasted about 50 min.

3. Results and discussion

3.1. Comparison of final- and running-estimate conditions

Mean initial (T0) and final (T24) ratings of the 13 contingencies in each of the two estimation conditions are shown in Table 1 along with the mean change scores (T24 − T0) in each case. In general, judgments sensitively reflected the overall Δp in effect, without there being any strong indication that initial interevent information exerted greater impact than did terminal interevent information. There was also little evidence to suggest that the final- and running-estimation procedures importantly influenced participants' final ratings.

A 2 (condition) × 2 (initial versus final judgment) × 13 (problem) mixed analysis of variance (ANOVA) with Condition as a between-subject factor and with initial versus final judgment and problem as within-subject factors was conducted on the participants' ratings. Mean values are shown in Table 1. Significant main effects of initial versus final judgment [$F(1, 98) = 5.42$, $p < .05$] and problem [$F(12, 1176) = 41.36$, $p < .001$] were obtained as were significant interactions of initial versus final judgment × problem [$F(12, 1176) = 59.85$, $p < .001$] and condition × initial versus final judgment × problem [$F(12, 1176) = 1.89$, $p < .05$].

In order to elucidate these two interactions, follow-up ANOVAs were conducted. A 2 (condition) × 13 (problem) mixed ANOVA on the initial ratings yielded no significant main effects or interactions. Thus, prior to the receipt of any contingency information, ratings were similar in both conditions for all problems. There was a small positive bias in the initial ratings, with mean scores falling between +0.08 and +1.54 (see Table 1). In both the running- and the final-judgment conditions, participants' initial ratings across all 13 problems differed significantly from zero [$t(49) = 3.72$, $p < .01$ and $t(49) = 5.00$, $p < .001$, respectively].

A 2 (condition) × 13 (problem) mixed ANOVA on participants' final ratings yielded only a significant main effect of problem [$F(12, 1176) = 79.79$, $p < .001$]. Final ratings closely accorded with the overall interevent contingency that participants were given. Mean ratings with $\Delta p = +.667$ ranged from +5.32 to +6.36; mean ratings with $\Delta p = +.333$ ranged from +2.90 to +4.50; mean ratings with $\Delta p = .000$ ranged from −0.42 to +1.24; mean ratings with $\Delta p = -.333$ ranged from −3.02 to −2.26; and mean ratings with $\Delta p = -.667$ ranged from −5.02 to −4.96. Furthermore, there was no clear indication of primacy. So, for example, with Problems 3, 6, 7, 8, and 9, final ratings were neither substantially more positive in Problems 6 and 7 (entailing an initially positive contingency) nor more negative in Problems 8 and 9 (entailing an initially negative contingency) than they were in Problem 3 (entailing a noncontingent relation throughout) (see Table 1). If there was any consistent tendency in these data, it was for final ratings in the running-estimate condition to show a (nonsignificant) trend toward recency, with ratings in Problems 6 and 7 more negative and ratings in Problem 9 more positive than in Problem 3 (see Table 1).

Finally, we conducted a 2 (condition) × 13 (problem) mixed ANOVA on the initial–final rating *change* scores. The significant problem main effect [$F(12, 1176) = 59.85$, $p < .001$] again testifies to the sensitivity of participants' ratings to the contingencies that they were given. Mean changes in ratings with $\Delta p = +.667$ ranged from +5.22 to +5.24; mean changes in ratings with $\Delta p = +.333$ ranged from +1.62 to +4.34; mean changes in ratings with $\Delta p = .000$ ranged from −1.40 to +1.00; mean changes in ratings with $\Delta p = -.333$ ranged from −3.74 to −2.84; and mean changes in ratings with $\Delta p = -.667$ ranged from −5.98 to −5.70. The significant condition × problem interaction [$F(12, 1176) = 1.89$, $p < .05$] did suggest differential responsiveness of participants in the different rating conditions to different interevent relations. Subsequent *t*-tests disclosed significant differences only on Problems 2 and 10. We therefore conclude that the different rating procedures exerted little notable effect on the participants' ratings.

3.2. Momentary ratings in the running-estimate condition

The mean momentary contingency ratings of participants in the running-estimate condition on each of the 13 problems is illustrated in Fig. 1. The five panels of this figure represent the five different groupings of problems that are shown in Table 1. Each panel of this figure and its associated ANOVA speak to different features of the data under consideration.

Fig. 1a depicts the five problems with unchanging contingencies throughout all 24 trials. Beginning just above zero on Trial 0 (the initial judgment prior to the receipt of any interevent information), ratings rapidly diverged in clear accord with the active contingencies. A 5 (problem) × 25 (trial) repeated-measures ANOVA yielded a significant main effect of problem [$F(4, 196) = 65.68$, $p < .001$] and a significant problem × trial interaction [$F(96, 4704) = 10.67$, $p < .001$].

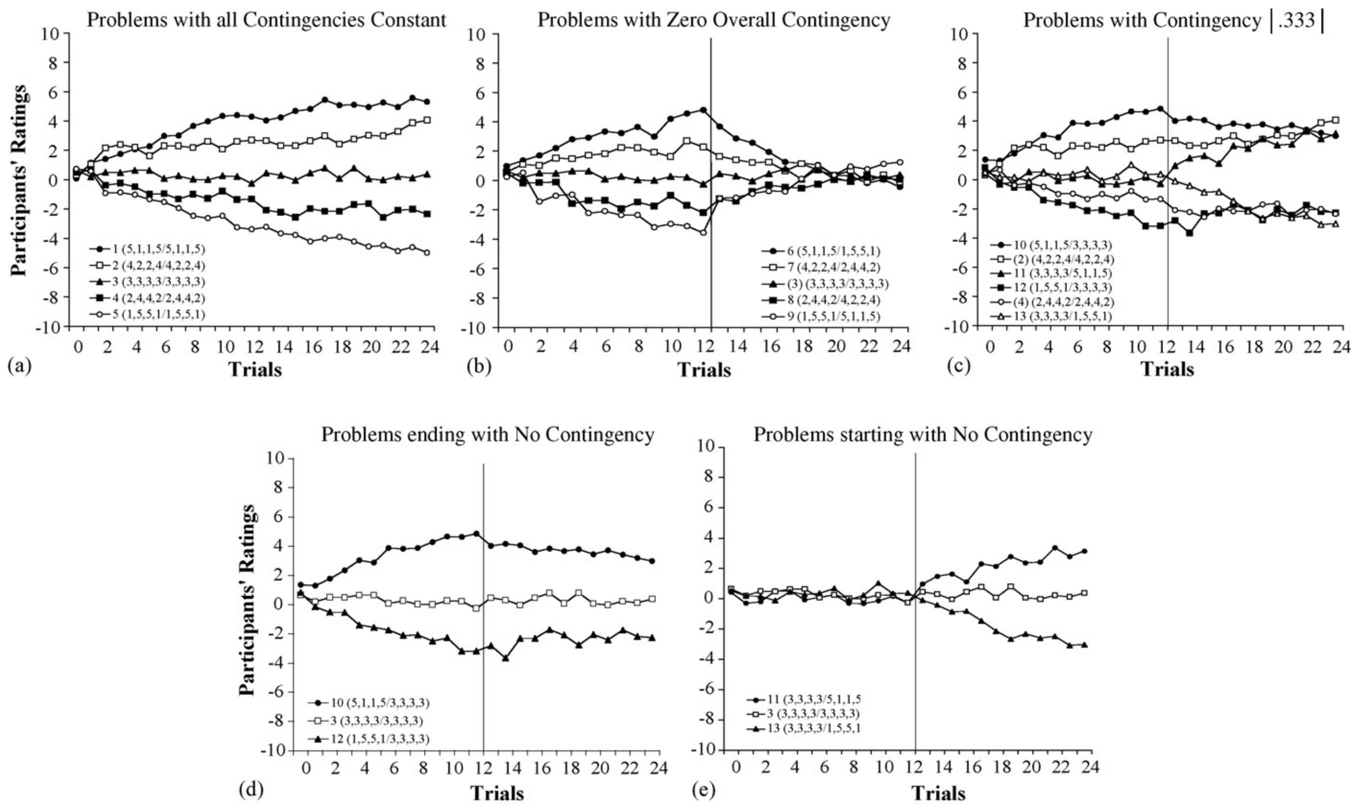


Fig. 1. Mean causal ratings of all 50 participants in the running-estimate condition in Experiment 1 across all trials on each of the 13 contingency problems. Panels a–e depict performance in the five sets of problems outlined in Table 1 and discussed at greater length in the text.

Fig. 1b depicts the five problems with no overall contingency across all 24 trials. As in Fig. 1a, ratings in the first 12 trials quickly diverged, in clear accord with the prevailing contingencies; however, ratings in the second 12 trials promptly converged, in synchrony with the changed interevent relations then in effect. A 5 (problem) × trial (25) repeated-measures ANOVA yielded a significant main effect of problem [$F(4, 196) = 22.76$, $p < .001$] and a significant problem × trial interaction [$F(96, 4704) = 9.20$, $p < .001$].

Fig. 1c depicts the six problems with an absolute overall Δp of .333. As in Fig. 1a and b, ratings in the first 12 trials rapidly diverged, in clear accord with the prevailing contingencies; however, ratings in the second 12 trials soon converged in the two subgroups of problems with overall Δps of +.333 and −.333, in synchrony with the changed interevent relations then in effect. A 6 (problem) × 25 (trial) repeated-measures ANOVA yielded a significant main effect of problem [$F(5, 245) = 52.42$, $p < .001$] and a significant problem × trial interaction [$F(120, 5880) = 8.81$, $p < .001$].

Fig. 1d illustrates the three problems entailing a noncontingent relation during the second 12 trials. After diverging during the first 12 trials, ratings during the second 12 trials converged, but not completely, in accord with the different overall contingencies in effect for these three problems. A 3 (problem) × 25 (trial) repeated-measures ANOVA yielded a significant main effect of problem [$F(2, 98) = 16.23$, $p < .001$] and a significant problem × trial interaction [$F(48, 2352) = 12.29$, $p < .001$].

Finally, Fig. 1e illustrates the three problems entailing a noncontingent relation during the first 12 trials. After remaining near zero during the first 12 trials, ratings during the second 12 trials promptly diverged, in accord with the changed contingencies then in effect. A 3 (problem) × 25 (trial) repeated-measures ANOVA yielded a significant main effect of problem [$F(2, 98) = 66.81$, $p < .001$] and a significant problem × trial interaction [$F(48, 2352) = 6.56$, $p < .001$].

These data thus disclose that our participants' ratings of interevent contingency not only sensitively accorded with the *overall* interevent relations to which they were exposed, but they closely tracked any *momentary* changes in those relations with little measurable lag. Just how might they accomplish such impressive cognitive feats?

We next examined whether an associative learning model such as that of Rescorla and Wagner (1972) could account for these contingency judgments. This model states that changes in the associative strength of a stimulus result from the outcome – reinforcement or nonreinforcement – on each trial involving the presentation of that stimulus. Specifically, this model states that the effectiveness of a target event $F(V_F)$ (e.g., the receipt of a fertilizer in our plant-fertilizer example) in producing the occurrence of a particular outcome event B (e.g., the plant's blooming) is determined by the associative strength of the compound of the target event and the background stimuli (V_{FX}) (e.g., the receipt of the fertilizer and water) and by the associative strength of the background stimuli alone (V_X) (e.g., the receipt of water alone or other contextual factors). Moreover, changes in the associative strengths of the target event (ΔV_F) and the background stimuli (ΔV_X) with the outcome are determined by the following conditions, representing the interevent information in Cells A, B, C, and D, respectively:

1. *Cell A*: If the compound of the target event (F) and the background (X) is followed by the outcome (B), then changes in the associative strengths of the respective components may be represented as Eq. (1a):

$$\Delta V_F = \alpha_F \beta_B(\lambda_B - V_{FX}) \quad \text{and} \quad \Delta V_X = \alpha_X \beta_B(\lambda_B - V_{FX}). \tag{1a}$$

2. *Cell B*: If the compound of the target event (F) and the background (X) is *not* followed by the outcome (B), then changes in the associative strengths of the respective components may be represented as Eq. (1b):

$$\Delta V_F = \alpha_F \beta_{NB}(\lambda_{NB} - V_{FX}) \quad \text{and}$$
$$\Delta V_X = \alpha_X \beta_{NB}(\lambda_{NB} - V_{FX}). \tag{1b}$$

3. *Cell C*: If the background (X) alone (i.e., the absence of the target event F) is followed by the outcome (B), then the change in the associative strength of X, ΔV_X, may be represented as Eq. (1c):

$$\Delta V_X = \alpha_X \beta_B(\lambda_B - V_X). \tag{1c}$$

4. *Cell D*: If the background (X) alone (i.e., the absence of the target event F) is *not* followed by the outcome (B), then the change in the associative strength of X, ΔV_X, may be represented as Eq. (1d):

$$\Delta V_X = \alpha_X \beta_{NB}(\lambda_{NB} - V_X). \tag{1d}$$

As can be seen in the above equations, there are three sets of parameters that affect the magnitude of the changes involved. The α values are learning rate parameters (stimulus saliences) associated with the component stimuli: α_F is for the target event (F) and α_X is for the background (X). The β values are learning rate parameters associated with the outcome (B): β_B is for the occurrence of the outcome and β_{NB} is for the nonoccurrence of the outcome. The λ values are the asymptotic levels of associative strength that can be supported by the outcome: λ_B is for the occurrence of the outcome and λ_{NB} is for the nonoccurrence of the outcome. The values of α and β are confined to the unit interval, $0 \leq \alpha, \beta \leq 1$. The λ parameter is normally assigned a value of 1 if the outcome is present and 0 if the outcome is absent.

Furthermore, the Rescorla–Wagner model assumes that the relation among the associative strengths of the target event (F), the background (X), and the compound of the target event and the background (FX) is linear:

$$V_F = V_X = V_{FX}. \tag{2}$$

Based on the above assumption, we obtain the associative strength of the target event (F) alone with the outcome (B) from Eq. (3):

$$V_F = V_{FX} - V_X. \tag{3}$$

Given the fact that participants generally use the 2 × 2 information in the order Cell A > Cell B > Cell C > Cell D, we expected the salience of the target event (α_F) to be greater than that of the background (α_X) and we expected the learning rate parameter associated with the occurrence of the outcome (β_B) to be greater than that associated with the nonoccurrence of the outcome (β_{NB}). In fact, we previously estimated the following specific values for these learning parameters in the present task: $\alpha_F = .34$, $\alpha_X = .14$, $\beta_B = .31$, and $\beta_{NB} = .29$ (Kao and Wasserman, 1993). Those values came from the same data set in which we estimated the weights of Cells A, B, C, and D to equal: .41, .31, .18, and .09 (also see Van Hamme and Wasserman, 1994; Wasserman et al., 1993, 1996).

For quantitative purposes, we used the same analytical scheme adopted by Kao (1993), which fits each individual participant's momentary contingency ratings with the best set of parameters in the Rescorla–Wagner model (α_F, α_X, β_B, and β_{NB}; see Eqs. (1a)–(1d)) estimated from *only* the initial and final ratings of all 13 problems (using the Hooke and Jeeves method described by Nash and Walker-Smith, 1987). The estimated parameters are put into the relevant equations and trial-by-trial performance is simulated in accord with the specific sequence of cell information that each individual participant was given on each individual problem. Then, those estimated trial-by-trial ratings for each problem are averaged across all participants and plotted along with the actual empirical data.

By way of illustration, Fig. 2 depicts the actual and predicted data for a subset of the problems with zero overall contingency: two of which changed (Problems 6 and 9) and one of which remained constant (Problem 3). The Rescorla–Wagner model closely fit these data. Such close fits were representative of all 13 contingency problems (see Wasserman et al., 1996, for the entire

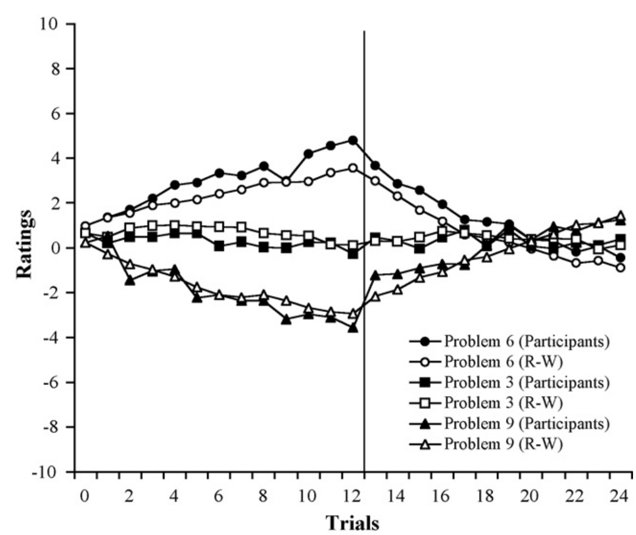


Fig. 2. Mean causal ratings of all 50 participants in the running-estimate condition in Experiment 1 on Problems 3, 6, and 9, plotted along with the ratings expected by Rescorla and Wagner's associative learning model.

set of simulations). Indeed, in our original analysis of these data, we found that the Rescorla–Wagner account better accorded with participants' momentary ratings than did Busemeyer's (1991) information integration model, which assigns different subjective weights to each of the four cells of the contingency table (see Wasserman et al., 1996).

4. Experiment 2

Having found in Experiment 1 that an associative account nicely accorded with participants' momentary and overall ratings of contingency, we decided to explore more precisely how such accord might arise. The specific issue that we investigated in Experiment 2 was the differential weighting of the four cells of the 2 × 2 contingency table.

Scrutiny of Table 1 discloses that all of the within-problem changes in contingency that we arranged in Experiment 1 were brought about by equivalent increases and decreases in the contents of each of the four cells of the 2 × 2 contingency table. So, in Problem 6, Cells A–D were each changed by 4; each of these changes (Cell A = −4, Cell B = +4, Cell C = +4, and Cell D = −4) was consistent with the change from a positive to a negative contingency. And, in Problem 10, Cells A–D were each changed by 2; each of these changes (Cell A = −2, Cell B = +2, Cell C = +2, and Cell D = −2) was consistent with the change from a positive contingency to a zero contingency. [Recall that entries in Cells A and D are consistent with positive interevent relations, whereas entries in Cells B and C are consistent with negative interevent relations.]

Although normative statistics, like Δp, equally weight those increments and decrements in the contents of the four cells of the 2 × 2 contingency table, human raters may not do so. If those nominally equivalent changes differentially affect participants' judgment behavior, then different combinations of changes in the contents of Cells A–D may have notably different effects on participants' causal ratings when contingencies are changed without warning in the middle of a problem. As we have seen, such differential effects are quite consistent with an associative account such as the Rescorla–Wagner model.

Our second experiment looked specifically for such differential effects of changing the contents of the 2 × 2 cell information. In order to appreciate the experimental strategy that we deployed, Table 2 lists all 14 problems that we gave to each participant. Problems 1 and 8 were constant positive and constant negative contingencies, respectively; they served as yardsticks for assessing the effects of the 12 changed contingencies. Problems 2 and 9 involved reversed contingencies, going from positive to negative relations and from negative to positive relations, respectively. The 10 remaining problems switched from contingent to noncontingent relations by changing the contents of all *four* cells or only *two* of the four cells of the 2 × 2 contingency table: five (Problems 3–7) went from a positive to a zero contingency and five (Problems 10–14) went from a negative to a zero contingency. The momentary and overall changes in Δp that these cell changes produced were virtually equal. However, based on prior work showing that the psychological impact of the four 2 × 2 contingency cells was ordered Cell A > Cell B > Cell C > Cell D, we expected that the five different ways of switching from contingency to noncontingency that we arranged would differ in a highly specific fashion: namely, *equally summed* changes in the contents of Cells A–D should produce changes in participants' ratings that are ordered Cells A and B > Cells A and C > Cells A–D > Cells B and D > Cells C and D (see Kao and Wasserman, 1993 for further discussion of the investigative tactic of making pairwise changes in the cells of the 2 × 2 contingency table).

To illustrate, Problems 3–7 all entail a combined numerical change of 8 in the contents of the four cells. But, the ordered impact of cell information – Cell A > Cell B > Cell C > Cell D – led us to expect that the *fall* in ratings that participants exhibit from the 12th to the 24th trials of these positive-to-noncontingency problems will be ordered Problem 3 > Problem 4 > Problem 7 > Problem 5 > Problem 6. In a parallel manner, we expected that the *rise* in ratings that participants exhibit from the 12th to the 24th trials of negative-to-noncontingency Problems 10–14 will be ordered Problem 10 > Problem 11 > Problem 14 > Problem 12 > Problem 13. The numerical changes in Cells A–D from the first 12 trials to the second 12 trials should produce those predicted orderings; those predicted orderings require only that the psychological significance of the four cells is ordered Cell A > Cell B > Cell C > Cell D. The predicted orderings do not depend on the size of the disparities between cells.

Because there was no important difference in the ratings of participants in the running-estimate and final-estimate conditions of Experiment 1 and because it was necessary here to measure the *change* in ratings observed on the last trial of the first half (Trial 12) and on the last trial of the second half (Trial 24) of the changed contingencies in order to compare those changes across problem types, participants in Experiment 2 made only running estimates of contingency.

4.1. Method

4.1.1. Participants

The participants were 50 (25 male and 25 female) students from the same courses as participants in Experiment 1.

4.1.2. Apparatus

The apparatus was the same as in Experiment 1.

4.1.3. Design

The experiment entailed 14 different contingency problems, shown in Table 2. As described earlier, these problems enabled us to make precise predictions concerning the effect of implementing an unsignaled shift from a contingent to a noncontingent interevent relation after the first 12 trials of a 24-trial problem.

4.1.4. Procedure

The procedure was virtually the same as that given to participants in the running-estimate condition of Experiment 1. The

Table 2
Problems used in Experiment 2 and participants' ratings

Problems	Changed cell and amount	Cell entries												Contingencies			Participants' ratings				Adjusted score
		Block 1				Block 2				Overall				Block 1	Block 2	Overall	T0	T12	T24	T24 − T12	
		A	B	C	D	A	B	C	D	A	B	C	D								
Problem with constant positive contingency																					
1	No change	5	1	1	5	5	1	1	5	10	2	2	10	0.667	0.667	0.667	1.50	4.54	5.76	1.22	
Problems staring with positive contingency																					
2	16[ABCD]	5	1	1	5	1	5	5	1	6	6	6	6	0.667	−0.667	0.000	1.16	4.68	0.92	−3.76	−4.98
3	8[AB]	5	1	1	5	1	5	1	5	6	6	2	10	0.667	0.000	0.333	0.36	4.40	2.42	−1.98	−3.20
4	8[AC]	5	1	1	5	1	1	5	5	6	2	6	10	0.667	0.000	0.375	0.94	4.84	3.04	−1.80	−3.02
5	8[BD]	5	1	1	5	5	5	1	1	10	6	2	6	0.667	0.000	0.375	1.08	4.64	3.66	−0.98	−2.20
6	8[CD]	5	1	1	5	5	1	5	1	10	2	6	6	0.667	0.000	0.333	0.88	4.58	4.58	0.00	−1.22
7	8[ABCD]	5	1	1	5	3	3	3	3	8	4	4	8	0.667	0.000	0.333	0.36	5.44	3.76	−1.68	−2.90
Problem with constant negative contingency																					
8	No change	1	5	5	1	1	5	5	1	2	10	10	2	−0.667	−0.667	−0.667	0.70	−3.56	−5.34	−1.78	
Problems starting with negative contingency																					
9	16[ABCD]	1	5	5	1	5	1	1	5	6	6	6	6	−0.667	0.667	0.000	0.44	−3.06	−0.32	2.74	4.52
10	8[AB]	1	5	5	1	5	1	5	1	6	6	10	2	−0.667	0.000	−0.333	1.02	−3.58	−1.56	2.02	3.80
11	8[AC]	1	5	5	1	5	5	1	1	6	10	6	2	−0.667	0.000	−0.375	0.52	−3.74	−2.80	0.94	2.72
12	8[BD]	1	5	5	1	1	1	5	5	2	6	10	6	−0.667	0.000	−0.375	0.86	−3.24	−2.66	0.58	2.36
13	8[CD]	1	5	5	1	1	5	1	5	2	10	6	6	−0.667	0.000	−0.333	1.02	−2.84	−3.00	−0.16	1.62
14	8[ABCD]	1	5	5	1	3	3	3	3	4	8	8	4	−0.667	0.000	−0.333	1.36	−2.86	−2.02	0.84	2.62

addition of one problem meant that experimental sessions lasted about 55 min.

4.2. Results and discussion

Mean ratings on Trial 0 (the initial estimate that was made without any contingency information), on Trial 12 (the last trial of the first half of the problems), and on Trial 24 (the last trial of the problems) of each of the 14 contingencies are shown in Table 2 along with the mean change scores from Trial 12 to Trial 24 in each case. In general, Trial 0 ratings were slightly positive, Trial 12 ratings clearly accorded with the positive and negative contingencies to which participants were exposed in the first half of the problems, and Trial 24 ratings clearly accorded with the overall contingencies to which participants were exposed. More importantly, changes in ratings from Trial 12 to Trial 24 differed depending on the specific cells of the 2×2 contingency table that were changed from the first to the second half of the problems.

A one-way repeated-measures ANOVA on the Trial 0 ratings of the 14 problems failed to yield a significant main effect. Thus, prior to the receipt of any contingency information, ratings were similar for all problems. There was a small, but reliable [$t(49) = 4.68$, $p < .001$] positive bias in the Trial 0 ratings, with participants' mean scores across all 14 problems falling between +0.36 and +1.50 (see Table 2).

A one-way repeated-measures ANOVA on the Trial 12 ratings of the 14 problems yielded a significant main effect [$F(13, 637) = 69.45$, $p < .001$]. Problems 1–7 involving $\Delta p = +.667$ had mean ratings ranging from +4.40 to +5.44, whereas Problems 8–14 involving $\Delta p = -.667$ had mean ratings ranging from −3.74 to −2.84 (see Table 2).

A one-way repeated-measures ANOVA on the Trial 24 ratings of the 14 problems yielded a significant main effect [$F(13, 637) = 40.02$, $p < .001$]. Problem 1 involving an overall $\Delta p = +.667$ had a mean rating of +5.76, Problems 3–7 involving overall Δps = +.333 and +.375 had mean ratings ranging from +2.42 to +4.58, Problems 2 and 9 involving an overall $\Delta p = .000$ had mean ratings ranging from −0.32 to +0.92, Problems 10–14 involving overall Δps = −.375 and −.333 had mean ratings ranging from −3.00 to −1.56, and Problem 8 involving an overall $\Delta p = -.667$ had a mean rating of −5.34 (see Table 2).

Raw rating change scores from Trial 12 to Trial 24 continued upward (change = +1.22) for Problem 1 involving a constant $\Delta p = +.667$ and downward (change = −1.78) for Problem 8 involving a constant $\Delta p = -.667$. Obviously, performance was not yet asymptotic on Trial 12. Therefore, the raw change scores between Trial 12 and Trial 24 for Problems 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, and 14 had to be adjusted in order to appreciate this fact.[1] Doing so was simple and direct. Each participant's (usually positive) Trial 24 − Trial 12 rating change score on Problem 1 was subtracted from each of the corresponding rating change scores on Problems 2–7; and each participant's (usually negative) Trial 24 − Trial 12 rating change score on Problem 8 was subtracted from each of the corresponding rating change scores on Problems 9–14. The means of those adjusted scores are also given in Table 2 and are graphically displayed in Fig. 3.

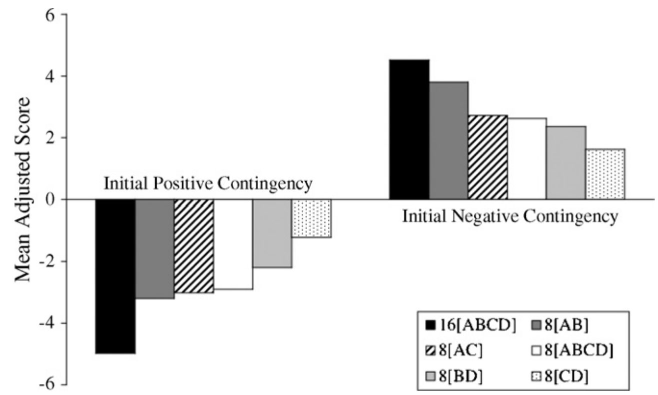


Fig. 3. Mean adjusted causal rating change scores of all 50 participants in Experiment 2 for those contingency problems entailing changes in Cells A–D equaling 8 and 16; for those problems entailing changes in Cells A and B equaling 8; for those problems entailing changes in Cells A and C equaling 8; for those problems entailing changes in Cells B and D equaling 8; and for those problems entailing changes in Cells C and D equaling 8. Performance is separately depicted for those problems involving initially positive contingencies and for those involving initially negative contingencies.

Reversing the relation from positive to negative (adjusted change score = −4.98) or from negative to positive (adjusted change score = +4.52) had larger effects on participants' ratings than did switching to a noncontingent relation from either a positive relation (adjusted change scores ranged from −1.22 to −3.20) or from a negative relation (adjusted change scores ranged from +1.62 to +3.80). More interestingly, switching to a noncontingent relation from either a positive or a negative relation had *different* effects on participants' ratings depending on the specific cells in the contingency table that were changed. For both initial positive and initial negative relations, the ordered impact of cell changes was: Cell A (4) and Cell B (4) > Cell A (4) and Cell C (4) > Cell A (2), Cell B (2), Cell C (2), and Cell D (2) > Cell B (4) and Cell D (4) > Cell C (4) and Cell D (4). This obtained ordering is inconsistent with participants' using the information from each cell equivalently, but it is consistent with participants' using the information in the order Cell A > Cell B > Cell C > Cell D.

A 2 (initial positive contingency versus initial negative contingency) × 6 (specific cell change) repeated-measures ANOVA was conducted on the adjusted Trial 24 − Trial 12 rating change scores. The Initial Contingency main effect was statistically significant [$F(1, 49) = 48.06$, $p < .001$] as was the Initial contingency × specific cell change interaction [$F(5, 245) = 9.61$, $p < .001$].

In order to elucidate this interaction, separate one-way [specific cell change (6)] ANOVAs were conducted on those

[1] Conclusions from the adjusted scores are just the same as conclusions without adjusting these scores; we used the adjusted scores to compensate for the fact that performance was not at asymptote at the beginning of the second half of training.

problems that began with positive relations and with negative relations. Each of these ANOVAs yielded statistically significant main effects of specific cell changes: for initially positive contingencies [$F(5, 245) = 4.87$, $p < .001$] and for initially negative contingencies [$F(5, 245) = 6.05$, $p < .001$]. Then, individual *t*-tests were conducted to find out which of these change scores differed significantly from zero. Of the 12 *t*-tests that were conducted, 11 were significantly different from zero; Problem 6 with changes = 4 in each of Cell C and Cell D fell just short of statistical significance [$t(49) = -1.95$, $p < .06$].

Although these statistical tests suggested that the effects of specific cell changes were similar for problems beginning with either positive relations or with negative relations, an additional repeated-measures ANOVA was conducted on the adjusted change scores, now recalibrated so that more extreme change scores were in each case reflected by more positive scores. This recalibration was done by multiplying all of the adjusted change scores in Problems 2–7 by −1 and by multiplying all of the adjusted change scores in Problems 9–14 by +1. [This simple process changed the mean adjusted change scores in Table 2 by reversing the sign of the scores in Problems 2–7 from − to +.] A 2 (initial positive contingency versus initial negative contingency) × 6 (specific cell change) repeated-measures ANOVA was conducted on the adjusted and recalibrated Trial 24 − Trial 12 rating change scores. The specific cell change main effect was statistically significant [$F(5, 245) = 9.61$, $p < .001$], but the Initial contingency × specific cell change interaction was not.

It was finally of interest to determine which specific cell changes differed reliably from one another. So, pairwise tests were conducted on the recalibrated rating change scores. Problems 2 and 9 combined (involving changes in each of Cells A–D = 4) differed reliably from Problems 4 and 11 combined (involving changes in each of Cells A and C = 4), $F(1, 49) = 10.95$, $p = .001$, Problems 5 and 12 combined (involving changes in each of Cells B and D = 4), $F(1, 49) = 15.99$, $p < .001$, Problems 6 and 13 combined (involving changes in each of Cells C and D = 4), $F(1, 49) = 29.77$, $p < .001$, and Problems 7 and 14 combined (involving changes in each of Cells A, B, C, and D = 2), $F(1, 49) = 14.77$, $p < .001$. Also, Problems 3 and 10 combined (involving changes in each of Cells A and B = 4) differed reliably from Problems 6 and 13 combined (involving changes in each of Cells C and D = 4), $F(1, 49) = 13.55$, $p < .001$.

5. General discussion

The present pair of experiments show that college students are extremely sensitive judges of interevent contingency. Their ratings of the contingent relations between dichotomous variables accorded very well with overall Δps ranging from −.667 to +.667. This accord between objective interevent relations and participants' psychological ratings is now well-established (for a review of this evidence, see Allan, 1993; Shanks and Dickinson, 1987; Wasserman, 1990; Wasserman et al., 1996; also see Jenkins and Ward, 1965 and Smedslund, 1963 for initial discouraging evidence); at least under optimal conditions, humans are quite able judges of the causal texture of the environment.[2]

Not only did this accord hold over the entirety of several different contingency problems, but it held when the relations within a single problem changed at the halfway point without warning. Most notably, those participants who made trial-by-trial ratings of interevent contingency sensitively tracked changes in the prevailing relations of several different sorts: (a) shifts from contingency to noncontingency, (b) shifts from noncontingency to contingency, and (c) reversals in contingency from positive to negative relations and from negative to positive relations. Finally, whether participants made trial-by-trial ratings of the prevailing relations or participants made only a final rating at the end of each problem, they showed no notable tendency for early (or late) interevent information to exert a disproportionate influence on their contingency ratings—reliable primacy (or recency) was not obtained.

Does this behavioral evidence therefore mean that participants were processing the contingency information that they were given as though they were computing some proper statistic of interevent association like Δp? Not necessarily. Here, as elsewhere (e.g., Catena et al., 1998; Kao and Wasserman, 1993; Levin et al., 1993; Mandel and Lehman, 1998; Schustack and Sternberg, 1981; Wasserman et al., 1990, 1993; White, 2003), the details of participants' rating behavior did not correspond with their use of a normative measure of contingency; participants disproportionately *weighted* the evidence in the contingency table in the order Cell A > Cell B > Cell C > Cell D. This differential weighting meant that when either a positive or a negative contingency in the first half of a problem was unexpectedly switched to noncontingency in the second half of a problem in Experiment 2, different ways of creating that noncontingency led to different magnitudes of change in participants' ratings. Specifically, *equally summed* changes in the contents of Cells A–D produced changes in participants' ratings that were ordered Cells A and B > Cells A and C > Cells A–D > Cells B and D > Cells C and D. According to normative statistical measures, however, all of these different methods should have had the same behavioral effect.

Rescorla and Wagner's (1972) associative learning theory (Eqs. (1a)–(1d))—succeeds in this aim by assigning differential salience to added stimulus events and to background stimuli as well as by assigning differential effectiveness to reinforcement and nonreinforcement: α_F, α_X, β_B, and β_{NB}, respectively, in the present problem setting. Of course, from these data alone, we cannot decide the ultimate utility of this model or the plausibility of other future accounts. We do note, however, that the Rescorla–Wagner model did not expect a pronounced primacy

[2] Even though we did not find any systematic bias in participants' ratings, there are other contingency and causal studies that show a number of deviations from the programmed contingencies. For example, it has often been observed that causal judgments are affected by the frequency of the effect, even when Δp is held constant (Allan and Jenkins, 1983; Baker et al., 2000; Lober and Shanks, 2000; Perales and Shanks, 2003). It has also been observed that causal judgments are affected by the frequency of the cause, even when Δp is held constant (Perales et al., 2005).

effect—and none was obtained. Kao (1993) too failed to obtain primacy in her study of contingency judgment. Only Yates and Curley (1986) and Dennis and Ahn (2001) found results suggestive of primacy. More recently, Stout et al. (2005) observed that, when testing occurred immediately after training, participants' predictive ratings were based on the most recent information; but, when testing occurred after a 48 h delay, primacy could be obtained. Further work into primacy in contingency judgments – both direct and derived – is certainly warranted.

With regard to the differential impact of each of the cells in the contingency table, Levin et al. (1993) suggested that people have a preference for positively stated information (as given in Cell A) and that *a priori* hypotheses and expectancies about the interevent relations being rated can determine which pieces of information are deemed to be more relevant (also see Klayman and Ha, 1987). Following this argument, White (2003) observed that contingency judgment tasks are often framed in such a way as to focus on the positive value of the cue and the outcome (Cell A). Also, Mandel and Lehman (1998) suggested that participants have a preference for *sufficiency* information over *necessity* information; they consider information as to whether a cause is sufficient to produce an effect (Cells A and B) to be more important than information as to whether the cause is necessary to produce the effect (Cells C and D).

A final topic for discussion centers on what some might believe to be the surprising effectiveness of an associative model to explain causal judgments (see Allan, 1993, for more extensive discussion of this issue). But, is it really so surprising that associative learning might effectively explain judgments of interevent contingency? Writing over 250 years ago, David Hume (1739/1964) argued that causal perception was reducible to associative learning. His own rules by which we come to form causal connections bear a strong resemblance to the principles of associative learning in animals that were so ably captured by the Rescorla–Wagner model.

Of course, the Rescorla–Wagner model of associative learning is hardly the last word on the topics of learning in animals or causal and contingency judgment in humans. Data in each domain exist that disclose important limitations on the model's applicability (see Allan, 1993, for more on this matter). One of those limitations is of particular interest to us here. Earlier, we discussed forward blocking as a key demonstration of a sequence effect in animal learning and human causal attribution, one that was quite nicely explained by the Rescorla–Wagner model. Here, we note that, at least in human judgment situations, there is good reason to believe that "backward blocking" is also a robust result (Chapman, 1991). In this case, after several pairings of the AX compound with the outcome, later pairings of Stimulus A alone with the outcome reduce the control previously acquired by Stimulus X. This backward blocking effect amounts to a form of retrospective revaluation that cannot be explained by the original Rescorla–Wagner formulation, because changes in associative strength were assumed to occur only on trials on which a particular stimulus is given. Training with Stimulus A alone after AX training should have no effect on Stimulus X, as it was never given during the second training stage.

Nevertheless, backward blocking and other potential embarrassments of the original Rescorla–Wagner model can be overcome by a simple modification in the model that assigns *nonzero* salience to *nonpresented* stimuli (Van Hamme and Wasserman, 1994). This revision may prove to be especially significant in extending an associative account to behavioral facts of a decidedly "cognitive" or "logical" character (Wasserman and Castro, 2005).

Acknowledgement

The participation of Masayoshi Katagiri in this research was made possible by the Japanese Ministry of Education.

References

Allan, L.G., 1980. A note on measurement of contingency between two binary variables in judgment tasks. Bull. Psychonomic Soc. 15, 147–149.

Allan, L.G., 1993. Human contingency judgments: rule based or associative? Psychol. Bull. 114, 435–448.

Allan, L.G., Jenkins, H.M., 1983. The effect of representations of binary variables on judgment of influence. Learning Motiv. 14, 381–405.

Arcediano, F., Matute, H., Miller, R.R., 1997. Blocking of Pavlovian conditioning in humans. Learning Motiv. 28, 188–199.

Arkes, H.R., Harkness, A.R., 1983. Estimates of contingency between two dichotomous variables. J. Exp. Psychol.: Gen. 112, 117–135.

Baker, A.G., Vallée-Tourangeau, F., Murphy, R.A., 2000. Asymptotic judgment of cause in a relative validity paradigm. Memory Cogn. 28, 466–479.

Busemeyer, J.R., 1991. Intuitive statistical estimation. In: Anderson, N.H. (Ed.), Contributions to Information Integration Theory, vol. I. Erlbaum, Hillsdale, NJ, pp. 187–215.

Catena, A., Maldonado, A., Cándido, A., 1998. The effect of frequency of judgement and the type of trials on covariation learning. J. Exp. Psychol.: Human Percept. Perform. 24, 481–495.

Chapman, G.B., 1991. Trial order affects cue interaction in contingency judgment. J. Exp. Psychol.: Learning Memory Cogn. 17, 837–854.

Chapman, G.B., Robbins, S.J., 1990. Cue interaction in human contingency judgment. Memory Cogn. 18, 537–545.

Church, R.M., 1964. Systematic effect of random error in the yoked control design. Psychol. Bull. 62, 122–131.

Church, R.M., 1969. Response suppression. In: Campbell, B.A., Church, R.M. (Eds.), Punishment and Aversive Behavior. Appleton-Century Crofts, Oxford, England, pp. 11–156.

Collins, D.J., Shanks, D.R., 2002. Momentary and integrative response strategies in causal judgment. Memory Cogn. 30, 1138–1147.

Dennis, M.J., Ahn, W.K., 2001. Primacy in causal strength judgments: the effect of initial evidence for generative versus inhibitory relationships. Memory Cogn. 29 (152), 164.

Dickinson, A., Shanks, D.R., Evenden, J., 1984. Judgement of act-outcome contingency: the role of selective attribution. Quart. J. Exp. Psychol. 36A, 29–50.

Hastie, R., Park, B., 1986. The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. Psychol. Rev. 93, 258–268.

Hume, D., 1739/1964. In: Selby-Bigge, L.A. (Ed.), Treatise of Human Nature. Oxford University Press, London.

Jenkins, H.M., Ward, W.C., 1965. The judgment of contingency between responses and outcomes. Psychol. Monogr. 79, 1–17.

Kamin, L.J., 1969. Selective association and conditioning. In: Mackintosh, N.J., Honig, W.K. (Eds.), Fundamental Issues in Associative Learning. Dalhousie University Press, Halifax, pp. 42–64.

Kao, S.-F., 1993. Information integration and associative learning as accounts of contingency judgment. Unpublished Doctoral Dissertation. University of Iowa.

Kao, S.-F., Wasserman, E.A., 1993. Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. J. Exp. Psychol.: Learning Memory Cogn. 19, 1363–1386.
Klayman, J., Ha, Y.W., 1987. Confirmation, disconfirmation, and information in hypothesis testing. Psychol. Rev. 94, 211–228.
Levin, I.P., Wasserman, E.A., Kao, S.-F., 1993. Multiple methods for examining the contributions of specific cell information to contingency judgments. Organ. Behav. Human Decis. Process. 55, 228–250.
Lober, K., Shanks, D.R., 1997. Is causal induction based on causal power? Critique of Cheng. Psychol. Rev. 107, 195–212.
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. J. Exp. Psychol.: Learning Memory Cogn. 24, 672–694.
Mandel, D.R., Lehman, D.R., 1998. Integration of contingency information in judgments of cause, covariation, and probability. J. Exp. Psychol.: Gen. 127, 269–285.
Matute, H., Vegas, S., De Marez, P.J., 2002. Flexible use of recent information in causal and predictive judgments. J. Exp. Psychol.: Learning Memory Cogn. 28, 714–725.
Miller, R.R., Matute, H., 1996. Biological significance in forward and backward blocking: resolution of a discrepancy between animal conditioning and human causal judgment. J. Exp. Psychol.: Gen. 125, 370–386.
Nash, J.C., Walker-Smith, M., 1987. Nonlinear Parameter Estimation: An Integrated System in BASIC. Marcel Dekker, New York.
Perales, J.C., Catena, A., Shanks, D.R., González, J.A., 2005. Dissociation between judgments and outcome-expectancy measures in covariation learning: a signal detection theory approach. J. Exp. Psychol.: Learning Memory Cogn. 31, 1105–1120.
Perales, J.C., Shanks, D.R., 2003. Normative and descriptive accounts of the influence of power and contingency on causal judgement. Quart. J. Exp. Psychol.: Human Exp. Psychol. 56A, 977–1007.
Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning. II. Current Theory and Research. Appleton-Century Crofts, New York, pp. 64–99.
Schustack, M.W., Sternberg, R.J., 1981. Evaluation of evidence in causal inference. J. Exp. Psychol.: Gen. 110, 101–120.
Shanks, D.R., 1985. Forward and backward blocking in human contingency judgment. Quart. J. Exp. Psychol. 37B, 1–21.
Shanks, D.R., Dickinson, A., 1987. Associative accounts of causality judgment. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation, vol. 21. Academic Press, San Diego, CA, pp. 229–261.
Smedslund, J., 1963. The concept of correlation in adults. Scand. J. Psychol. 4, 165–173.
Stout, S.C., Amundson, J.C., Miller, R.R., 2005. Trial order and retention interval in human predictive judgment. Mem. Cogn. 33, 1368–1376.
Van Hamme, L.J., Wasserman, E.A., 1994. Cue competition in causality judgments: the role of nonpresentation of compound stimulus elements. Learning Motiv. 25, 127–151.
Wasserman, E.A., 1988. Response bias in the yoked control procedure. Behav. Brain Sci. 11, 477–478.
Wasserman, E.A., 1990. Detecting response-outcome relations: toward an understanding of the causal texture of the environment. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation, vol. 26. Academic Press, New York, pp. 27–82.
Wasserman, E.A., Castro, L., 2005. Surprise and change: variations in the strength of present and absent cues in causal learning. Learning Behav. 33, 131–146.
Wasserman, E.A., Dorner, W.W., Kao, S.-F., 1990. Contributions of specific cell information to judgments of interevent contingency. J. Exp. Psychol.: Learning Memory Cogn. 16, 509–521.
Wasserman, E.A., Elek, S.M., Chatlosh, D.L., Baker, A.G., 1993. Rating causal relations: role of probability in judgments of response-outcome contingency. J. Exp. Psychol.: Learning Memory Cogn. 19, 174–188.
Wasserman, E.A., Kao, S.-F., Van Hamme, L.J., Katagiri, M., Young, M.E., 1996. Causation and association. In: Shanks, D.R., Holyoak, K., Medin, D.L. (Eds.), Causal Learning. Academic Press, San Diego, CA, pp. 207–264.
White, P.A., 2003. Effects of wording and stimulus format on the use of contingency information in causal judgment. Memory Cogn. 31, 231–242.
Yates, J.F., Curley, S.P., 1986. Contingency judgment: primacy effects and attention decrement. Acta Psychologica 62, 293–302.