Continuous monitoring of human contingency judgment across trials

DAVID R. SHANKS
University of Cambridge, Cambridge, England

Recent evidence suggests that people are sensitive to the degree of contingency between their actions and ensuing outcomes, but little is known about the way in which such contingency judgments develop as more and more information about the contingency is provided. Three experiments examined this issue in the context of a video game. In Experiment 1, it was found that contingency judgments follow growth functions: When the contingency was positive, judgments increased toward a positive asymptote, and when the contingency was negative, judgments decreased toward a negative asymptote. When the contingency was zero, judgments themselves remained close to zero but were biased by the overall frequency with which the outcome occurred. In Experiment 2, it was shown that the growth function was not the result of the anchoring of early judgments at the zero point. The bias in judgments when the contingency is zero was investigated in Experiment 3. The results are discussed in terms of rule-based analyses and contemporary theories of conditioning.

Causal relationships between our actions and their outcomes have an essential feature, namely, that there is a degree of contingency or temporal correlation between them. It is not surprising, therefore, that people have been found to be sensitive to this property. Allan and Jenkins (1980) and Alloy and Abramson (1979), for example, found that when the probability of an outcome (O) given an action (A), P(O/A), was held constant, judgments of the extent to which the action controlled the outcome decreased as the probability of the outcome in the absence of the action, P(O/−A), was increased. The result of increasing P(O/−A) is to decrease the contingency.

This sensitivity to the degree of act-outcome contingency has prompted some authors to examine which of several rules of judgment best accounts for their subjects' actual judgments. For instance, Allan and Jenkins (1980) found that the judgments of the majority of their subjects were most highly correlated with one of the normative measures of contingency (Allan, 1980), the ΔP rule, where

$$\Delta P = P(O/A) - P(O/-A).$$

A possible implication of this is that the subjects were estimating P(O/A) and P(O/−A) separately and arriving at their judgments by taking the difference between these terms. Allan and Jenkins also examined four other rules. In the ΔF rule, judgments are based on the difference be-

tween the frequency of outcomes given the action (O/A) and the frequency in the absence of the action (O/−A), whereas in the ΔD rule, judgments are based on the difference between the number of "confirming" instances (O/A and −O/−A) and the number of "disconfirming" instances (O/−A and −O/A). The #O rule bases judgments on the total number of outcomes (O/A and O/−A). Finally, in the #RO rule, judgments are based simply on the number of times the outcome and the action occur together (O/A).

In their experiments, Allen and Jenkins (1980) found that for each of the five strategies there was at least one subject whose judgment pattern was most highly correlated with that strategy, although the ΔP rule predominated. In a more recent report (Allan & Jenkins, 1983), however, these authors provided evidence that the ΔD rule, and not the ΔP rule, gave the best account of the overall pattern of contingency judgments.

Wasserman, Chatlosh, and Neunaber (1983) analyzed the same five rules. These authors calculated the proportion of subjects whose patterns of judgments were significantly correlated with a given rule. In this analysis, it was possible for a subject's judgments to be correlated with more than one rule. Although the ΔP rule was again found to predominate, Wasserman et al. reported that over half of their subjects' judgments were correlated significantly with the ΔD and ΔF rules.

A conclusion that can be drawn from these "rule-based" analyses of human contingency judgment is that a particular strategy, most likely the ΔP or ΔD rule, is used by subjects to determine their judgments of contingency. For example, Allan and Jenkins (1983) stated that "judgments tend to be based on the difference between confirming and disconfirming cases" (p. 381), whereas Wasserman et al. (1983) reported that their results "supported the conclusion that subjects rated response-outcome

158

relations on the basis of the difference in the probability of an outcome given their having recently made or not made a response" (p. 406).

Recently, however, a quite different view of contingency judgment has emerged. As some authors (e.g., Alloy & Abramson, 1979; Alloy & Tabachnik, 1984) have noted, there is a remarkable similarity between the factors that seem to affect human contingency judgment and those affecting animal Pavlovian and instrumental conditioning. In fact, on the basis of this similarity, it has been proposed that there exists a mechanism common to animals and humans for the detection of causal relationships (Alloy & Tabachnik, 1984). The results of human contingency judgment experiments and animal conditioning studies can be compared if the action is taken as being analogous to the conditioned stimulus (CS) in a Pavlovian procedure or the operant in an instrumental procedure, and the outcome as analogous to the unconditioned stimulus (US) or reinforcer. It is then assumed that the strength of an animal's conditioned response reflects its judgment about the degree of contingency between the conditioned stimulus or operant and the reinforcer.

There are three main phenomena common to the results of animal conditioning and human contingency judgment experiments. First, if the probability of the US in the presence of the CS, P(US/CS), is held constant, then the degree of excitatory conditioning decreases systematically as the probability of the US in the absence of the CS, P(US/−CS), is increased (Rescorla, 1968), just as human contingency judgments decrease as P(O/−A) is raised when P(O/A) is held constant (Alloy & Abramson, 1979; Allan & Jenkins, 1980). Secondly, if the action and the outcome are in fact noncontingent, judgments are related to the frequency of the outcome in certain cases (Alloy & Abramson, 1979; Dickinson, Shanks, & Evenden, 1984; however, Wasserman et al., 1983, failed to find this effect). Similarly, excitatory conditioning in animals is often found to increase with the frequency of the US during limited exposure to a noncontingent CS-US relationship (e.g., Rescorla, 1972). Finally, humans have been shown to be sensitive to negative action-outcome contingencies in which the action actually prevents the outcome from occurring (Dickinson et al., 1984, Experiment 1); in Pavlovian conditioning, a stimulus becomes an inhibitor, capable of opposing the action of an excitatory stimulus, if it has been established in a negative contingency with the US, when the probability of the US is greater in the absence of the CS than in its presence (Rescorla, 1969; Witcher & Ayres, 1980). These similarities encourage the view that contemporary theories of animal conditioning might be extended to human contingency judgment.

Dickinson et al. (1984) specifically attempted to test this possibility. Subjects in their experiment were each given six sets of trials presented in the context of a video game. Subjects were able to fire shells at a tank that moved across the video screen, so that the shells were one potential cause of tank destruction. An alternative cause was a minefield

through which the tank had to pass. In three of the conditions, the probability of tank destruction given that a shell was fired, P(O/A), was .75. The conditions differed in the probability of tank destruction in the absence of a shell, P(O/−A), which is equivalent to the probability of the minefield alone destroying the tank. This took the values .75, .5, and .25. The pattern of judgments was consistent with the results of Allan and Jenkins (1980) and Alloy and Abramson (1979): As P(O/−A) was increased, judgments of the effectiveness of the shells in destroying the tanks decreased. In three other conditions, P(O/−A) was held constant at .75, and P(O/A) was varied between .75, .5, and .25. As P(O/A) was decreased, judgments became more negative, matching the result often found in animal Pavlovian conditioning when a stimulus becomes an inhibitor, capable of opposing the effect of an excitor (Rescorla, 1969).

For two of the above conditions, the action and the outcome were noncontingent: In one case, P(O/A) and P(O/−A) were both .75; in the other case, they were both .25. Comparison of the judgments in these two cases replicated the finding of Alloy and Abramson (1979), that judgments increased as the frequency of the outcome was increased, despite the fact that the contingency was in fact zero.

Dickinson et al. (1984) then performed computer simulations based on two of the best-known contemporary theories of conditioning to see whether these theories could produce terminal judgments similar to those actually found. The two theories were those of Pearce and Hall (1980) and Rescorla and Wagner (1972). Although these theories were initially presented as accounts of Pavlovian conditioning, they can be regarded as general statements of the associative learning processes underlying both Pavlovian and instrumental conditioning (Dickinson, 1980). In each case, a set of parameters was found such that the simulation closely matched the experimental results. Each of the three main phenomena was reproduced: The decrease in judgments as P(O/−A) was increased with P(O/A) held constant; the increase in negative judgments as P(O/A) was decreased, with P(O/−A) held constant; and the greater judgments in the noncontingent set with the greater frequency of outcomes.

Although these conditioning theories are successful in explaining the pattern of judgments, it is not necessarily the case that subjects were arriving at their judgments in the way that the theories of conditioning would maintain. It is equally possible for the rule-based analysis to explain the pattern of results. For example, the results were entirely consistent with the ΔP rule, except for the finding that subjects were biased in the noncontingent sets to give judgments that depended on the frequency of the outcome. But it would be quite reasonable to have as an auxiliary to the ΔP rule the strategy that if ΔP is zero, then the number of outcomes should be taken into account. Obviously, the results of the Dickinson et al. (1984) experiment do not distinguish between these possibilities.

But despite the extent of our knowledge about judgments of contingency given after a series of trials involving an action and an outcome, almost nothing is known about the way in which such judgments develop during the course of the trials as more and more information about the contingency is provided. In the experiment of Dickinson et al.(1984) described above, for example, subjects were merely asked for their contingency judgments after 40 trials. Clearly, data on this issue should make it easier to distinguish between the two explanations of contingency judgment that have been described above. The purpose of the three experiments reported here was to provide such data.

EXPERIMENT 1

In the three experiments to be reported here, the video game used by Dickinson et al. (1984) was modified to allow for the monitoring of contingency judgments as they developed. Dickinson et al. (1984, Experiment 1) gave their subjects sets consisting of 40 "trials" or chances to fire at a tank and asked them to provide a judgment of contingency at the end of each set. For these experiments, however, the game was modified such that a judgment was given every 5 trials. The judgments were entered by moving a pointer on a linear scale that was constantly present on the screen, so that the subjects could see what the previous judgment had been.

In Experiment 1, the subjects were each given four sets of trials, with the action-outcome contingency varying across the sets. In one set, there was a positive contingency ($\Delta P = .5$), and in another set, a negative contingency ($\Delta P = -.5$). For the two remaining sets, ΔP was zero. These sets differed, however, in the probability of the outcome on each trial: For one set, this was .75, whereas for the other, it was .25.

Method
Subjects. The subjects were 22 students of both sexes who were specially recruited for the experiment and who were tested individually in an unfurnished experimental room. The subjects were asked to read carefully the instructions on the video screen, and told that any queries they had about the task would be answered. At the end of the session, the subjects were paid for their participation.

Apparatus. The game was run on an Acorn System 3 microcomputer. The computer was connected to a monochrome video display (screen size 19 cm wide and 14 cm high), placed approximately 30 cm in front of the subject. The subjects were required to respond to the instructions presented on the screen by typing on a computer keyboard placed immediately in front of them.

Procedure. At the beginning of the experiment, the subjects were given the following instructions on the video screen (responses that the subjects were required to make are given in parentheses):

The computer will take you through the procedure step by step. To move to the next step, press the SPACE-BAR. To move back to the previous step press the RETURN key. If you have any problems please ask for assistance.
READY?? (space-bar)
You have been given the task of testing a new kind of shell under battle conditions. The shells are to be fired at tanks which will pass across your gun-sights. Since the shells are new, some of

them may hit their target but fail to explode. You must try to evaluate the effectiveness of the shells.
(space-bar)
On the right of your screen is your sight. Tanks will pass across the screen from right to left. After the third, press the SPACE-BAR.

On the right of the screen the sight was displayed. This consisted of a rectangle, 4 cm wide and 3 cm high, centered 5 cm from the right-hand side of the screen and midway between the top and bottom of the screen. At random intervals between 0.8 and 3.2 sec (average = 2.0 sec), three representations of a tank, approximately 2 cm long and 1 cm high, appeared from the right edge of the screen and passed along a horizontal path through the sight and across the screen to its left-hand edge before disappearing. This intertrial interval was maintained throughout the experiment. The tank took 1.0 sec to cross the screen.
The next instruction was:

When they blow up, it looks like this.

A tank then appeared and "exploded" after passing through the sight. This destruction consisted of the fragmentation of the tank and then the disappearance of the remains of the tank. The point at which the tank was fragmented on its trajectory across the screen was determined randomly. The next instruction was:

You can shoot by pressing the SPACE-BAR as the tanks pass in front of the sight. Have a try!

A tank then appeared, and if the subject pressed the space-bar while the tank was within the sight, the tank exploded at some random point on its traverse. The tank did not explode immediately so that the destructions following a hit and those following no hit would be identical. Throughout the experiment, if a tank was programmed to explode on a particular trial, then the point at which it exploded was always random, regardless of whether a correctly aimed shot was fired or not. In other words, there was no systematic difference between where a tank was destroyed if it was hit by a shell and where it was destroyed if it hit a mine.
If the subject missed the first tank, more tanks appeared, until a successful shot had been registered. The next instruction, when the subject had fired on target, was:

You may have noticed that if your shot is on target, a blip appears above your sight.

A marker appeared throughout the experiment whenever a shot was on target. This marker was a small rectangle situated above the top right-hand corner of the sight. In order for the shell to hit the tank, the space-bar had to be pressed between 0.16 and 0.28 sec after the appearance of the tank, while it was in the gun-sight. This reaction time requirement was included so that the subject's attention would be maintained throughout the task. Since the subject could fire (press the space-bar) but miss the target, the term "hit" will be used to describe correctly aimed shots. Failure to score a hit means either that the subject fired and missed or did not fire at all.
A second set of practice trials was presented, and when the subject had again fired on target, the main instructions appeared:

It is your job to determine how good the shells are in blowing up the tanks. Of course, you can only evaluate a shell's effectiveness if the marker shows that your shot hit the tank. Although the marker will come up instantly for a correctly aimed shot, the shot will not hit the tank for some time since it is a long distance away. The tank will have entered a minefield before the shot has reached it, so the destruction can be due either to the shell or to a mine. (space-bar)
A correctly aimed shot may hit the tank anywhere within the

minefield. Thus, it is impossible to tell from a single trial with a correctly aimed shot whether the tank exploded because it hit a mine or because it was hit by your shell. Of course, if you do not fire, or your aim was incorrect, and the tank explodes, it must be because it hit a mine. (space-bar)

Since the shells are new, some of them may hit the tank but fail to explode. So only a proportion of correctly aimed shots will cause the tank to blow up. In addition, being struck by a an ineffective shell may alert the tank driver, so that he drives more carefully through the minefield. Thus a shot may actually reduce the likelihood of the tank being destroyed.

(space-bar)

Every few trials you will be asked to give your judgment of the effectiveness of the shells on a scale from −100 to +100. Positive scores indicate that firing a shell increased the likelihood of the tank being destroyed and negative scores that firing decreased the likelihood. Zero indicates that the shells were completely ineffective. To enter your judgment, move the pointer left on the scale below by pressing 1 and right by pressing 0. Have a try!

A scale was presented at the bottom of the screen, going from -100 on the left to +100 on the right, and calibrated in units of 5. Below the scale was a small arrow that served as a pointer. The subjects were able to move the pointer left by pressing 1 on the keyboard and right by pressing 0.

The instructions continued when the subject had seen how to move the pointer:

There will be four sets altogether, each consisting of 40 chances to fire. The effectiveness of the shells and the density of the minefield may well differ from one set to the next. At the start of each set you will be asked to type in a number indicating where you would like the pointer to start on the scale. Since you will not have experienced the shells or minefield for that set, you will have no evidence on which to base a judgment, but just type in whatever number you wish the scale to start at. The first set will begin when you press the SPACE-BAR.

(space-bar)

Type in a number below to indicate your judgment of the effectiveness of the shells. Press the RETURN key after typing in the number.

The subjects were then required to type in a number between −100 and +100 to indicate their initial bias. Having done this, the scale appeared at the bottom of the screen, with the pointer indicating that number. The subjects were asked to type in a number at the beginning of each set. When the subject pressed the space-bar, the game itself began. The scale remained on the screen throughout the set of trials, with the pointer marking the subject's previous judgment, and after every fifth trial the subject was asked to modify his or her judgment:

Move the marker to indicate your judgment of the effectiveness of the shells. Press 1 to move left and 0 to move right.

In one set (75-25), there was a positive contingency: P(O/A) = .75 and P(O/−A = .25, and therefore ΔP = .5. In a second set (25-75), the contingency was negative (ΔP = −.5): P(O/A) = .25 and P(O/−A) = .75. In the remaining sets (75-75 and 25-25), the contingency was zero: for one set, P(O/A) = P(O/−A) = .75, and for the other set, P(O/A) = P(O/−A) = .25. The order in which the sets were presented was random. Because the destruction or nondestruction of the tank on a particular trial was governed by a random number procedure in the program, the actual value of ΔP for each subject usually differed slightly from the nominal value. In the analysis, however, the actual contingency experienced by the subject was calculated, cumulated over all of the trials.

Results and Discussion

The data were analyzed by two separate analyses of var-

iance. The judgments of effectiveness, actual contingency, and number of hits were subjected to a 4 (contingency) × 8 (blocks of trials) analysis. Both of these factors are within subjects: Contingency distinguishes the four sets that each subject received, and blocks of trials refers to judgments given after every five trials. In addition, the initial bias scores were subjected to a one-way analysis, with four levels of contingency.

The mean initial bias scores are shown on the left of Figure 1, at the zero point on the trials axis. The analysis revealed that there was no effect of contingency (F < 1). The initial bias scores, collapsed across contingency, did not differ from zero (t < 1).

Over the 40 trials of each of the conditions 75-25, 75-75, 25-25, and 25-75, the mean numbers of hits were, respectively, 25.3, 23.5, 25.1, and 23.8. There were no significant differences among these scores (F < 1).

Figure 1 shows the mean cumulative actual values of ΔP. These were close to their nominal values, as Figure 1 illustrates. There was no difference in the values of ΔP across trials (F < 1), and no trials × contingency interaction [F(21,408) = 1.02, p > .25].

Figure 1 also shows the mean judgments of effectiveness for each condition across trials. There was a significant difference between the conditions [F(3,63) = 33.39, p < .001] and a trials × contingency interaction [F(21,441) = 1.84, p < .05]. The main effect of trials was not significant [F(7,147) = 1.33, p > .1]. Judgments in the different conditions were compared by three planned, nonorthogonal contrasts using the Bonferroni procedure to control the experimentwise error rate (Miller, 1966). With this procedure, α is divided by the number of contrasts to prevent the inflation of the Type I error rate. In each case, the comparison was based on
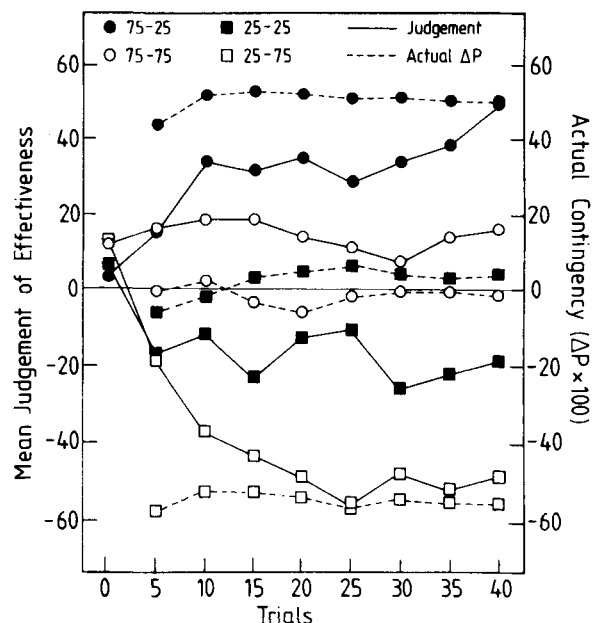

Figure 1. Mean actual contingencies (ΔP × 100) and mean judgments of contingency across trials for each of the conditions in Experiment 1.

the last two judgments in the set, in other words, after 35 and 40 trials, in order to provide sufficient data. There were significant differences between the 75-25 and 75-75 conditions [$F(1,21) = 12.19, p < .01$], the 75-75 and 25-25 conditions [$F(1,21) = 13.49, p < .01$], and the 25-25 and 25-75 conditions [$F(1,21) = 7.72, p < .05$].

Judgments after 5 and 40 trials were compared for each condition. For the noncontingent sets (75-75 and 25-25), these judgments did not differ ($t < 1$ in each case). For the 75-25 set, however, judgments were significantly greater after 40 than after 5 trials [$t(21) = 3.47, p < .01$], whereas for the 25-75 set, they were lower [$t(21) = 3.10, p < .01$, two-tailed].

The pattern of terminal judgments in this experiment is consistent with previous reports (e.g., Dickinson et al., 1984) that subjects are sensitive to different degrees of action-outcome contingency. However, the primary aim of the experiment was to provide some data about the way in which such judgments develop over the course of a series of experiences involving the action and the outcome. The results clearly imply that judgments follow a growth function that reaches asymptote at approximately the actual degree of contingency. As can be seen in the 75-25 and 25-75 conditions, in which there was a degree of contingency between the action and the outcome, judgments started from close to the zero point and moved incrementally or decrementally away toward an asymptote. But before the implications of these results for the various mechanisms of contingency judgment that have been proposed are considered, a possible trivial explanation of the growth function will be analyzed.

EXPERIMENT 2

In Experiment 1, the subjects were asked to give an initial bias score at the beginning of each set of trials. The means of these scores were all approximately zero, despite the fact that there was no particular reason for the subjects to enter one number rather than another, and this was taken as evidence that the subjects did not begin each set with any particular bias or expectation about the forthcoming action-outcome contingency. However, it is known that in many judgment situations an initial judgment will tend to bias or "anchor" subsequent judgments (Tversky & Kahneman, 1974). In Experiment 1, it could be argued that starting from the zero point would tend to anchor judgments such that they would only slowly move away from zero. This would then give results similar to those of a growth function.

In order to rule out an "anchoring" effect produced by the explicit request for an initial bias score, the subjects were not asked for such a score in Experiment 2.

Method

Subjects and Apparatus. The subjects were 28 students who participated during an undergraduate psychology practical class. Three Acorn System 3 microcomputers were placed adjacent to one another on a table in the practical classroom, and the subjects performed the task as part of the class. Except where specifically mentioned, the general experimental method was identical to that of Experiment 1.

Procedure. The subjects were given the same preliminary training and instructions as in Experiment 1, except for one change in the instructions. In the instructions employed in Experiment 1, it was never explicitly stated that the subjects should base their judgments at any particular point on all of the trials up to that point, although the judgments suggest that this is what they were doing in general. However, to guard against the possibility that a particular subject might base each judgment only on the past five trials, the final set of instructions was altered to the following:

> There will be four sets altogether, each set consisting of 40 chances to fire. The effectiveness of the shells and the density of the minefield will be constant within each set but may well vary from one set to another. Each judgment should be based not just on the preceding few trials but on all the trials so far of that set. The first set will begin when you press the SPACE-BAR.

When the subject pressed the space-bar, the game itself began. For the first five trials of the set, the scale did not appear on the screen. After the fifth trial, the subjects were asked for their judgments:

> Type in a number below to indicate your judgment of the effectiveness of the shells. Press the RETURN key after typing in the number.

The scale then appeared at the bottom of the screen, with the pointer marking on the scale the number that the subject had typed in. The scale remained on the screen during the rest of the set.

As in Experiment 1, there were four sets, presented in a random order. The parameters for the four sets were as before, 75-25, 75-75, 25-25, and 25-75, where the first number refers to $P(O/A) \times 100$ and the second to $P(O/-A) \times 100$. Again, the actual contingencies differed slightly from the nominal ones.

Results

The data for the number of hits, the cumulative actual contingency ΔP, and judgments of effectiveness were analyzed by means of 4 (conditions) × 8 (blocks of trials) analyses of variance.

There were no significant differences between the mean number of hits in each condition. For Conditions 75-25, 75-75, 25-25, and 25-75, the mean numbers of hits over the 40 trials were 24.4, 22.4, 23.4, and 22.9, respectively, with no significant differences between the conditions ($F < 1$). The conditions × trials interaction was not significant [$F(21,567) = 1.21, p > .1$].

The mean cumulative actual contingency, ΔP, for each condition is shown in Figure 2. As can be seen in Figure 2, the contingencies were close to their nominal values by the 10th trial. The actual values of ΔP did not vary across trials ($F < 1$), and there was no trials × condition interaction [$F(21,518) = 1.04, p > .25$].

Figure 2 also shows the mean judgments of effectiveness. There was a significant main effect of conditions [$F(3,81) = 41.76, p < .001$], a significant effect of trials [$F(7,189) = 11.06, p < .001$], and a significant conditions × trials interaction [$F(21,567) = 4.73, p < .001$]. Judgments after 35 and 40 trials were combined, and the same set of preplanned contrasts that was used in Experiment 1 compared these judgments in the different conditions. Judgments were found to be significantly different
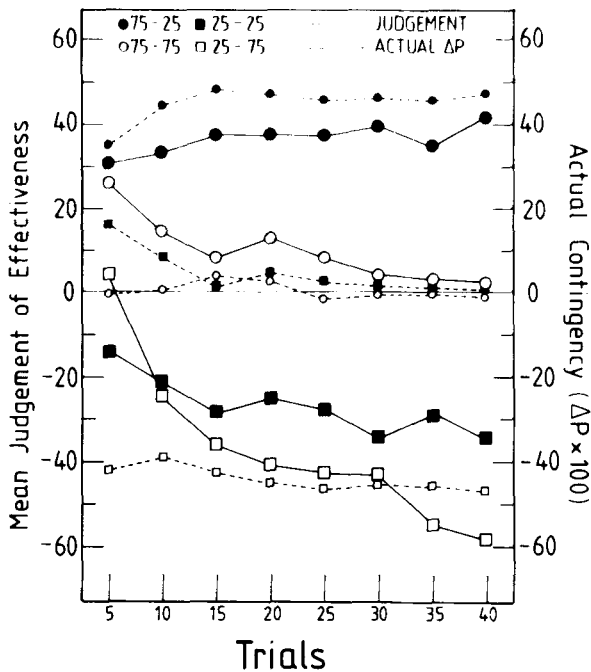
Figure 2. Mean actual contingencies (ΔP $\times$ 100) and mean judgments of contingency across trials for each of the conditions in Experiment 2.

for each of the comparisons: Between the 75-25 and 75-75 conditions, $F(1,27) = 16.91$, $p < .01$; between the 75-75 and 25-25 conditions, $F(1,27) = 9.03$, $p < .05$; and between the 25-25 and 25-75 conditions, $F(1,27) = 11.71$, $p < .01$.

Judgments of effectiveness for each particular condition were also analyzed. These were found to be significantly lower after 40 than after 5 trials for Condition 75-75 [$t(27) = 2.51$, $p < .05$], Condition 25-25 [$t(27) = 2.12$, $p < .05$], and Condition 25-75 [$t(27) = 7.92$, $p < .001$, two-tailed]. For Condition 75-25, judgments were significantly greater after 40 than after 5 trials, but only by a one-tailed test [$t(27) = 1.79$].

Discussion

These results confirm the finding of Experiment 1 that contingency judgments follow growth functions, since again there was a significant effect of trials and a trials $\times$ condition interaction. The principal outcome of Experiment 2, however, is that it demonstrates that the growth pattern persists when the possibility of anchoring brought about by an initial judgment is removed, although the growth of judgments in the 75-25 condition is clearly not as great as that seen in the equivalent set in Experiment 1. The results differed from those of Experiment 1, however, in that judgments in the two noncontingent sets decreased across trials. A possible explanation of this is that in Experiment 1, judgments in these conditions were anchored close to zero at the beginning of the set.

Any explanation of contingency judgment obviously has to account for the growth functions found in these two

experiments. One simple theory from the field of Pavlovian conditioning is capable of doing just this. The best-known contemporary theory of conditioning, that of Rescorla and Wagner (1972), proposes an equation for changing by increments or decrements the associative strength of a stimulus, which is assumed to determine the strength of the conditioned response it elicits. The course of learning is described by cumulating these changes across a series of trials. The theory states that the associative strength (V) of a CS will be changed on a trial according to the equation

$$\Delta V = \alpha \cdot \beta(\lambda - \overline{V}),$$

where ΔV is the change in associative strength, α is a learning rate parameter for the particular CS, β is a learning rate parameter for the particular US, λ is the asymptote of conditioning that the US is capable of supporting, and $\overline{V}$ is the total associative strength of all the conditioned stimuli present on that trial. For human contingency judgment, the associative strength of the action for the outcome is assumed to determine a subject's judgment of the extent to which the action is effective in producing the outcome.

In order to show that this theory can account for the results of Experiments 1 and 2, computer simulations based on it were run. Five hundred sets of judgments were derived according to the theory for each of the conditions, attempting to match as closely as possible the simulation to the actual conditions of the experiments. Each simulation produced a judgment every 5 trials, up to a total of 40 trials, based on the same set of instructions for determining the outcome of each trial that was used in the experiments. The same four conditions (75-25, 75-75, 25-25, and 25-75) were used. The probability of a hit on a given trial was set at .5; this produces equal sampling of the two conditional probabilities and is close to the actual figure of .60, based on all the subjects of Experiments 1 and 2.

Figure 3 shows that the simulation can produce a set of judgments that closely matches the actual data. The parameters used in the simulation were as follows: α for a hit, 0.7; α for the minefield, 0.6; β for the destruction of a tank, 0.3; β for the nondestruction of a tank, 0.6; and λ, 100. The top panel of Figure 3 illustrates the combined data of Experiments 1 and 2. Although it is possible to find sets of parameters that generate simulations to match each of the experiments separately, the data of these two experiments were put together in order to smooth out the curves. The bottom panel of Figure 3 shows the results of the simulation. It is evident that the simulation reproduces most of the salient features of the actual data. There is an increase across trials in the 75-25 set and a decrease in the 25-75 set, both curves tending toward asymptotic values. In the simulation, there is also a difference between the noncontingent sets, and a decrease in judgments across trials in these conditions. It should be pointed out that in the simulation the associa-
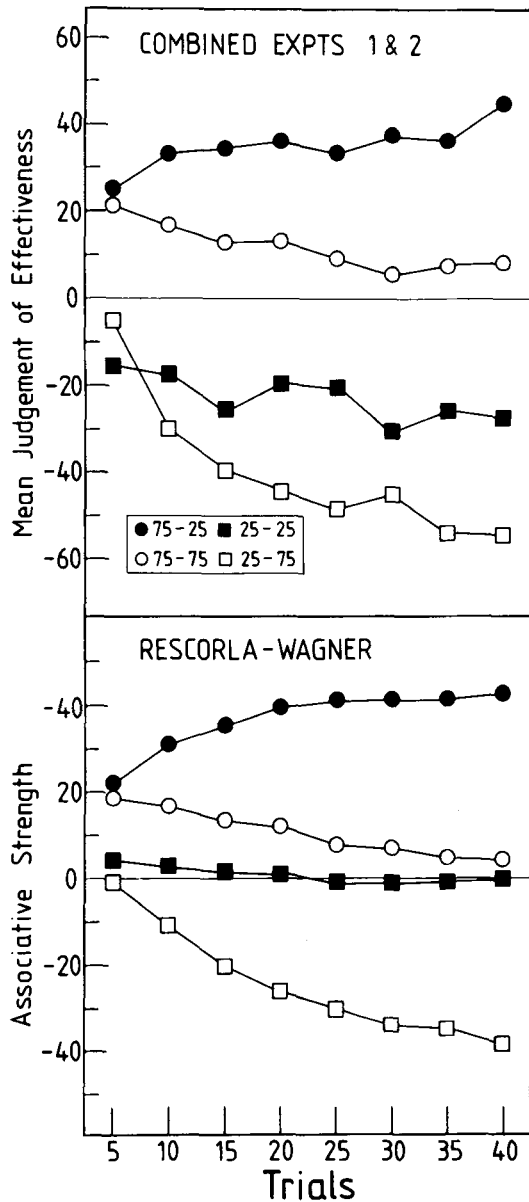
Figure 3. Mean judgments of contingency across trials for each condition, for Experiments 1 and 2 combined (top panel), and the results of a computer simulation (bottom panel) based on the Rescorla-Wagner model. The parameters employed in the simulation were: α for a hit, 0.7; α for the minefield, 0.6; β for the destruction of a tank, 0.3; β for the nondestruction of a tank, 0.6; and λ, 100. In each case, the associative strengths of a hit and of the minefield were assumed to start at zero.

tive strength always begins from zero: Therefore, in the 75-75 set, judgments start at zero, go up to a value of approximately 20 by the fifth trial, and then slowly decrease with more trials.

There is one obvious discrepancy between the simulation and the experimental results, however, in that the simulation gives judgments in the 25-25 condition that reach asymptote at zero, whereas the experimental results indicate an asymptote of around -30. In fact, the whole

25-25 curve seems to be displaced. Although there is no reason why there should be a strict correspondence between the associative strength of the action as predicted by the simulation and the judgment produced, and likewise no reason to suppose that the zero points of these two dimensions should coincide, the theory does predict that the asymptote in both of these conditions should be the same (zero). The issue of the difference between judgments in the 75-75 and 25-25 sets is addressed in Experiment 3.

Although the Rescorla-Wagner model accounts quite well for the results of the first two experiments, it is still quite possible for one or more of the rule-based analyses to do so also. There are problems, however, with some of the rules. For example, since the actual contingency remained constant across trials, the ΔP rule has trouble explaining the change in judgments across trials. Although the variance of ΔP should decrease as the sample size, and hence the number of trials, on which it is based increases, its mean should remain constant, and so, therefore, should a judgment based on it. In addition, the ΔP rule cannot explain the difference between judgments in the 75-75 and 25-25 conditions, in each of which ΔP is zero. The ΔF rule, which bases judgments on the difference between the number of outcomes given the action and outcomes given no action, has difficulty in accounting for the fact that judgments in both the noncontingent sets of Experiment 2 decreased across trials. In fact, this corroborates an experiment by Dickinson et al. (1984, Experiment 2), in which judgments were found to be significantly lower after 60 trials than after 30 trials in the 75-75 condition. In noncontingent sets, ΔF is always zero, since the number of outcomes occurring with the action is the same as the number occurring without it. Similarly, the ΔD rule, in which judgments are specified by the difference between the number of confirming and disconfirming instances, predicts judgments to be independent of trials in noncontingent sets.

EXPERIMENT 3

A salient feature of both of the previous experiments, a finding that has been noted in several other reports (e.g., Alloy & Abramson, 1979; Allan & Jenkins, 1980, 1983; Dickinson et al., 1984) is that judgments in the noncontingent sets were found to be dependent on the overall frequency with which the outcome occurred. Because this effect is not invariably found (e.g., Wasserman et al., 1983), it remains a possibility that it is not fundamentally due to the event frequency, as conditioning theories suggest (Dickinson et al., 1984), but is an artifact of the experimental procedure used.

One such possible artifact is easy to identify. In contrast with most Pavlovian conditioning experiments, in which each animal typically experiences only one degree of contingency between the CS and the US, each subject in the first two experiments witnessed four action-outcome contingencies over a short period of time. In these cir-

cumstances, it would be surprising if there were no interaction between the different sets. This suggests a possible explanation for the effect of the frequency of the outcome on judgments of contingency in noncontingent sets, that the subjects' impression of the density of the minefield carries over from one set to the next. In fact, this seems quite a plausible notion, because although it is emphasized in the instructions that the density of the minefield and the effectiveness of the shells may well differ from one set to the next, the subjects' attention is focused more on the shells than on the minefield. The average frequency of tank destruction by the minefield alone in the 75-25, 25-25, and 25-75 sets is 42%, which is less than the frequency in the 75-75 condition. Simply on the basis of the actual contingency, therefore, subjects should give positive judgments in a 75-75 set if an impression of the density of the minefield in preceding sets is carried over. The reverse is true for the 25-25 set: The average frequency of destruction by the minefield in the 75-25, 25-75, and 75-75 sets is 58%, which is considerably greater than the frequency in the 25-25 set. Again, on the basis of the actual contingency, subjects would be expected to give negative judgments in this set if knowledge about the minefield is carried over from preceding sets.

Experiment 3 attempted to assess this possibility. The subjects were again presented with four sets of trials. As before, in one of the sets, there was a positive contingency, and in another, a negative contingency. The other two sets were identical, either 75-75 for one group of subjects or 25-25 for another group. Each subject experienced the same noncontingent set after a set with a positive ΔP and after one with a negative ΔP. If the difference between judgments in the two types of noncontingent set is due to an impression of the density of the minefield in the preceding sets being carried over, then we would expect judgments to be greater when the preceding set was 75-25 than when it was 25-75. The possibility of such a carryover effect was assessed by asking subjects for an initial bias score before each set. If, however, it is the rate of outcomes that produces the difference, then we would expect no difference between judgments when the preceding set is 75-25 and when it is 25-75.

Method

Subjects and Apparatus. The subjects were 24 students who were tested individually in the same conditions and using the same apparatus as in Experiment 1. There were paid at the end of the session for their participation.

Procedure. Except where specifically mentioned, all details of the experiment were the same as in Experiment 1.

The subjects were given the same preliminary training and instructions as before. The instructions describing the game and the subjects' task were unchanged from Experiment 2, and the subjects were also required to enter an initial bias score at the beginning of each set.

Four sets were presented, as before. However, in this experiment, each subject received only one of the noncontingent sets, which occurred twice, in different "contexts": once after a set with a positive contingency (75-25) and once after a set with a negative contingency (25-75). For half of the subjects (Group 75-75), the

noncontingent set was 75-75; for the other half (Group 25-25), it was 25-25. In addition, the order in which the 75-25 and 25-75 sets were presented was counterbalanced. Thus, half of the subjects received sets in the order 75-25, noncontingent, 25-75, noncontingent, and the other half received them in the order 25-75, noncontingent, 75-25, noncontingent.

Results and Discussion

The results from the 75-25 and 25-75 sets were similar to those of the previous experiments and will not be discussed further.

The data from the noncontingent sets were analyzed by means of a $2 \times 2 \times 2 \times 8$ ANOVA, with the factors of group (75-75 vs. 25-25), the order of the sets, the context (following either a 75-25 or a 25-75 set), and trials, and by a separate 2×2 ANOVA for the initial bias scores, with the two factors being group and context.

On the left of each panel of Figure 4 are shown the mean initial bias scores. Analysis of variance revealed no effect of context (whether the previous set was 75-25 or 25-75) (F < 1) or of group (F < 1), but there was a significant group × context interaction [F(1,10) = 6.29, p < .05]. Subsequent t tests using the error term derived from the ANOVA found a significant difference between judgments in Group 25-25 [t(11) = 2.40, p < .05, two-tailed], but not between judgments in Group 75-75 [t(11) = 1.40, p > .10]. Given the fact that the set immediately preceding the initial bias was the same for subjects in the two groups, this interaction is difficult to interpret. However, it is obvious, at least for Group 25-25, that the subjects were biased as a result of the preceding set.

Figure 4 illustrates the mean judgments of effectiveness and the actual contingency ΔP across the 40 trials. The data are presented separately for Groups 25-25 (left-hand panel) and 75-75 (right-hand panel).

The mean numbers of hits out of 40 trials in Groups 75-75 and 25-25 were, respectively, 23.1 and 24.1. There was no difference between these scores and no groups × trials interaction (both Fs < 1). As Figure 4 shows, there were slight differences between the actual and nominal values of ΔP. There was no difference between the two groups (F < 1) and no effect of context [F(1,10) = 1.29, p > .25], but there was a significant effect of trials [F(7,70) = 2.73, p < .05]. None of the interactions were significant (Fs < 1).

Figure 4 also illustrates the mean judgments of effectiveness. Judgments were greater in Group 75-75 than in Group 25-25, in accordance with the results of the previous experiments [F(1,10) = 5.18, p < .05], but there was no effect of trials [F(7,70) = 1.88, p > .05] and no groups × trials interaction [F(7,70) = 1.36, p > .1]. The result of most interest, however, is that there was no effect of context on the judgments [F(1,10) = 2.90, p > .1] and no significant interactions involving this factor [largest F(7,77) = 1.21, p > .25]. Although there appears to have been an effect of context on judgments in Group 75-75, this difference was due largely to one subject and was not significant. In addition, it was in the
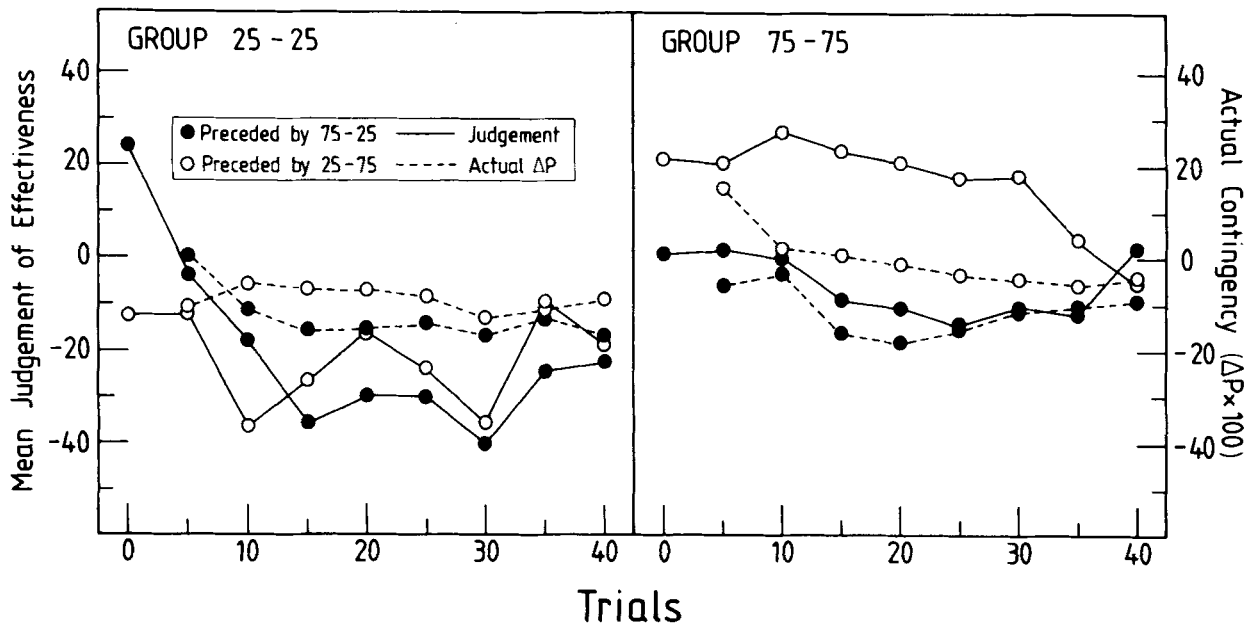
Figure 4. Mean actual contingencies ($\Delta P \times 100$) and mean judgments of contingency across trials for each of the conditions in Experiment 3. The left-hand panel illustrates the data for Group 25-25, and the right-hand panel the data for Group 75-75.

opposite direction to what would be predicted if an impression of the effectiveness of the minefield were carried over from one set to the next.

These results indicate, therefore, that judgments are unaffected by any impression of the density of the minefield in preceding sets that might have been carried over. Even though there was an effect on the initial bias scores, at least for Group 25-25, such that a preceding 75-25 set engendered a greater expectancy of action-outcome contingency than a preceding 25-75 set, judgments themselves were unaffected. This in turn implies that the difference between judgments in the 75-75 and 25-25 sets is not due to carryover effects. The fact that Alloy and Abramson (1979) found an equivalent difference in a between-subjects design strengthens this conclusion. An alternative explanation for this difference, which is the explanation offered by conditioning theories, is that it is a result of the frequency with which the outcome occurs.

GENERAL DISCUSSION

The aim of these experiments has been to examine the way in which judgments of action-outcome contingency develop as more and more information about the contingency is provided. In Experiment 1, it was found that judgments followed growth functions. When the contingency was positive, judgments incremented toward a value similar to the actual contingency, whereas when the contingency was negative, judgments were decremented across trials, again toward a value similar to the actual contingency. When the contingency was zero, judgments did not change across trials, but were dependent on the overall frequency of the outcome. Initial judgments given

at the beginning of each set indicated that the subjects began each set with no particular bias about the contingency.

In Experiment 2, it was found that judgments still followed growth functions when the possibility of the anchoring of early judgments close to zero was removed. However, it was found that when subjects were not asked for an initial bias score, judgments in the noncontingent sets decreased across trials.

Experiment 3 examined the origin of the dependence of judgments in the noncontingent sets upon the overall frequency of the outcome. A possible explanation was considered, that some impression about the density of the minefield in preceding sets might be carried over. It was found that judgments were unaffected by whether the preceding set had been one with a positive or a negative contingency, and it was therefore concluded that any impression of the density of the minefield in preceding sets is insignificant.

Researchers interested in contingency judgments have proposed two types of mechanism by which they might be generated. Rule-based analyses suggest that people use simple rules by which information can be combined to produce contingency judgments. Conditioning theories offer an alternative explanation. The results of Experiments 1 and 2 imply that a successful theory of contingency judgment, in addition to accommodating the known features of terminal judgments of contingency, must be able to account for the fact that judgments follow growth functions. Whereas it is difficult, for example, for the ΔP rule to do this, a simple theory of Pavlovian conditioning, the Rescorla-Wagner model, predicts just such growth functions. The simulation based on this model (Figure 3) shows that conditioning theory should be con-

sidered a serious candidate for explaining human contingency judgment. All of the principal features of the results of Experiments 1 and 2 are reproduced in the simulation: the increase in judgments in the 75-25 set and the decrease in the 25-75 set across trials; the difference between judgments in the two noncontingent sets; and the decrease in judgments across trials in the noncontingent sets that was seen in Experiment 2. The experimental results do not fulfill, however, the prediction of the Rescorla-Wagner model that judgments in the two noncontingent sets should asymptote at the same point. Experiment 3 suggests that the difference in judgments in the two noncontingent sets can be accounted for by the frequency of the outcome in the two conditions.

In summary, these data provide considerable evidence that the case for the application of conditioning theories to account for human contingency judgment is strong enough to merit serious consideration.

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.

ALLAN, L.G., & JENKINS, H.M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology*, 34, 1-11.

ALLAN, L.G., & JENKINS, H.M.(1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, 14, 381-405.

ALLOY, L.B., & ABRAMSON, L.Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, 441-485.

ALLOY, L.B., & TABACHNIK, N.(1984). Assessment of covariation by

humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, 91, 112-149.

DICKINSON, A. (1980). *Contemporary animal learning theory.* Cambridge: Cambridge University Press.

DICKINSON, A., SHANKS, D. R., & EVENDEN, J. L. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarerly Journal of Experimental Psychology*, 36A, 29-50.

MILLER, R. G. (1966). *Simultaneous statistical inference.* New York: McGraw-Hill.

PEARCE, J. M., & HALL, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.

RESCORLA, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology.* 66, 1-5.

RESCORLA, R. A. (1969). Conditioned inhibition of fear resulting from negative CS-US contingencies. *Journal of Comparative and Physiological Psychology*, 67, 504-509.

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

RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A .H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research.* New York: Appleton-Century-Crofts.

TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

WASSERMAN, E. A., CHATLOSH, D. L., & NEUNABER, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learning and Motivation*, 14, 406-432.

WITCHER, E. S., & AYRES, J. J. B. (1980). Systematic manipulation of CS-US pairings in negative CS-US correlation procedures in rats. *Animal Learning and Behavior*, 8, 67-74.