

Psychological Monographs: General and Applied

JUDGMENT OF CONTINGENCY BETWEEN RESPONSES
AND OUTCOMES¹

HERBERT M. JENKINS

AND

WILLIAM C. WARD

*McMaster University**Duke University*

3 experiments are reported in which Ss were asked to judge the degree of contingency between responses and outcomes. They were exposed to 60 trials on which a choice between 2 responses was followed by 1 of 2 possible outcomes. Each S judged both contingent and noncontingent problems. Some Ss actually made response choices while others simply viewed the events. Judgments were made by Ss who attempted to produce a single favorable outcome or, on the other hand, to control the occurrence of two neutral outcomes. In all conditions the amount of contingency judged was correlated with the number of successful trials, but was entirely unrelated to the actual degree of contingency. Accuracy of judgment was not improved by pretraining Ss on selected examples, even though it was possible to remove the correlation between judgment and successes by means of an appropriate selection of pretraining problems. The relation between everyday judgments of causal relations and the present experiment is considered.

AN important part of human verbal knowledge about the everyday physical and social environment is knowledge about what causes what. No doubt much of that knowledge is acquired from others and entails an understanding, at various levels of detail, of how the relation between a particular cause and its effect is mediated. We know about the relation between the setting of a thermostat and the temperature of a house, not as a result of raw observation, but through our understanding of the relation of thermostat to furnace and of furnace to heat. On the other hand, some knowledge about cause and effect sequences, whether valid or not, must arise primarily from the individual's experience with the way things happen. One may come to believe that wet weather is the cause of various bodily ills even though one has little prior notion of how such a relation might be mediated.

How are causes identified from experience? There is no difficulty in identifying a cause when consequent follows antecedent quickly and regularly. The relation between

the movements of a steering wheel and the behavior of a car, or between the flick of a switch and the appearance of a light are quickly perceived. But causes are also identified on the basis of less determinate observations. Thus, one may decide that a remark made yesterday caused someone to change his behavior today, or that taking a drug produced recovery from an illness. It is clearly more difficult to correctly identify a causal relation in cases of this type. The increased difficulty arises, at least in part, from the fact that the outcome occurs with some frequency in the absence of the antecedent in question (e.g., recoveries sometimes occur without drugs); and the antecedents are sometimes present when the outcomes are not (taking a drug is not always followed by recovery).

In the simple cases where the perception of a relation is immediate, the joint occurrence of two events stands out against a background of experience in which neither event has appeared alone with appreciable frequency. A single joint occurrence may, in such cases, lead to the conviction that the events are causally related. In the less determinate or noisier cases, however, the joint occurrence of antecedent and consequent does not have the same force. If antecedent and consequent each occur without

¹The experiments were carried out while the authors were members of staff of the Bell Telephone Laboratories, Murray Hill, New Jersey. The support of this research by the laboratories is gratefully acknowledged.

the other, their joint occurrence can arise through chance as well as through the result of causal relation. Thus the problem becomes one of estimating whether the frequency of joint occurrence exceeds what might be expected by chance. The estimate must rest upon a series of observations. When dealing with imperfectly related events of this sort, it seems more appropriate to speak of the judgment of a causal relation rather than of its perception.

The present experiments were designed to yield some preliminary information on how accurately people judge the degree of relation between events when the actual dependency is varied from zero (independent events) to some intermediate degree well short of a determinate or completely dependent relation. They were also concerned with the basis of the judgments. The situation, in brief, was this. The subjects were given two response buttons with which they tried to influence the appearance of two outcomes. On each of a series of trials they chose to press one of the response buttons and were then shown the outcome which followed. At the end of the series of trials they were asked to judge the degree of *control* which their response choices had exerted over the outcomes. We used the term "control" rather than alternative terms such as "dependency" or "correlation" because in the context of the task it seemed to be the most natural way to communicate the technical meaning of contingency with everyday language.

It will be useful to have an index of the actual degree of control or contingency between response choices and outcomes. The basic meaning of control is that the outcome depends upon the response. More exactly, there is control when the probability of a particular outcome given one response is different from the probability of that outcome given another response. The magnitude of the difference in these conditional outcome probabilities provides a simple index of the amount of control. For the present case with response alternatives R_1 and R_2 and outcomes O_1 and O_2 , the index of contingency, ΔP , is given by: $|\Pr(O_1/R_1) - \Pr(O_1/R_2)|$. The expression $\Pr(O_1/R_1)$ is read, the probability of O_1 given R_1 . The

range of values of ΔP is from one (complete control) to zero (no control). It is zero when the probability of O_1 given R_1 is the same as the probability of O_1 given R_2 . It is one when O_1 always follows given R_1 and never follows given R_2 , or vice versa. The value of the index is unchanged if the conditional probabilities for O_2 are used in place of those for O_1 since $P(O_2) = 1 - P(O_1)$.

If the four possible response-outcome pairs are arranged in a double entry (2×2) table with cells labeled $a = R_1, O_1$; $b = R_1, O_2$; $c = R_2, O_1$; $d = R_2, O_2$, the ΔP index is given by

$$\left| \frac{a}{a+b} - \frac{c}{c+d} \right|,$$

which simplifies to

$$\left| \frac{ad - bc}{(a+b)(c+d)} \right|.$$

Two experiments, one by Inhelder and Piaget (1958) and one by Smedslund (1963), are directly relevant to the present problem.

Inhelder and Piaget examined the concept of correlation in children of about 10-15 years of age. The children were shown a number of cards each with a face drawn on it. The faces had blue or brown eyes and blonde or brown hair. Each subject was asked about the relation of eye color to hair color for each of several different sets of cards. If the four possible pairings of eye color with hair color are arranged in a 2×2 table with cells labeled as follows: $a =$ blue eyes and blonde hair, $b =$ blue eyes and brown hair, $c =$ brown eyes and blonde hair, $d =$ brown eyes and brown hair, then the a and d cases, which make up one diagonal, are considered to be the confirming cases, while the b and c cases on the other diagonal are the nonconfirming cases.² The child was said to be using an explicit notion of correlation if his answers were based on the difference between the number of confirming and the number of nonconfirming cases in the set.

² In the absence of a specific hypothesis, the confirming cases are considered to be those on the diagonal having the larger total, whether $a + d$ or $b + c$.

Two stages in the child's approach are distinguished. In the first the child may organize the pictures into the four pairings; may talk about the chances of having, for example, blonde hair if you have blue eyes; and may identify confirming and nonconfirming cases. However, two features of the concept of correlation are missing. The first is that the a and d cases are not seen as equivalent and combined into one total, nor are the b and c cases taken together. Some children who do combine the cases properly, however, run into a second difficulty when they fail to relate the $a + d$ cases to the $b + c$ cases. In the more advanced stage of thinking these difficulties are overcome, and the child spontaneously relates confirming cases to nonconfirming cases and judges correlation in terms of the balance between the two. Concerning the proportion of children reaching this stage, the authors state only,

It is usually toward 14-15 years that the frequency of these cases is high enough to define a stage.

These results provide grounds, although not strong grounds, for expecting that adults are capable of making appropriate judgments of contingency in the present experiment. The grounds are weak on two counts. First, the data were displayed in quite a different manner. The instances upon which the judgment was based were small in number, they were all in view at one time, and they could be arranged by the subject into groups corresponding to the four types of pairings. In the present experiment, on the other hand, the instances are produced by the subject over an extended series of trials. A second, and more basic difference is that the logic of the concept of contingency as formulated by Inhelder and Piaget is less generally applicable than is the logic entailed by the ΔP index. The difference between the sum of the confirming cases and the sum of the nonconfirming cases can serve as an index of contingency only if the two states of at least one of the variables appear equally often. Otherwise, the sums may differ even though the variables are independent. For example, consider a set of instances in which eye color and hair color are in fact independent, but blue eyes

predominate over brown eyes, and blonde hair predominates over brown hair. For a particular example, let $a = 8$, $b = 2$, $c = 4$, $d = 1$, where, as before, the letters stand for frequencies of the joint occurrence of each of the four possible pairings. Here, there are more confirming cases ($a + d$) than nonconfirming cases ($b + c$). The ΔP index, however, is zero since the probability of having blonde hair given blue eyes, $\frac{8}{10}$, is not different from the probability of having blonde hair given brown eyes, or $\frac{4}{6}$. (The difference in the formulations can also be appreciated by noting that the numerator of the expression for ΔP in terms of cell frequencies is the difference between the *products* of the cell frequencies on the diagonals rather than a difference in their sums).

One cannot tell whether the successful subjects judged the correlation in terms of proportions or frequencies. In a number of the protocols given by Inhelder and Piaget the children do talk about "chances" or proportions rather than raw frequencies, and the authors make the general point that the concept of probability develops before that of correlation. However, in none of the cases reported were subjects presented with sets of instances containing disproportionate frequencies in the states of both variables.

The format of Smedslund's experiment was more similar to that of the present experiment. The subjects, who were nurses, attempted to judge the connection between a symptom and a diagnosis. On each of a series of cards a set of letters representing symptoms appeared together with another set of letters representing diagnoses. The attention of the subjects was directed toward whether or not a connection existed between one particular symptom and one particular diagnosis. These data can also be cast into a 2×2 table. The cells of the table contain the frequencies of the four possible pairings: presence or absence of the symptom with presence or absence of the diagnosis.

The judgments obtained by Smedslund (1963) showed no relation to the actual contingency between symptom and diagnosis. There was a substantial correlation

between just the frequency with which symptom and diagnosis appeared together (positive confirming cases only) and the number of subjects who thought that symptom and diagnosis were related. Smedslund concluded,

normal adults with no training in statistics do not have a cognitive structure isomorphic with correlation.

The effects of noncontingent reinforcement on performance in learning tasks has been studied. Although it is not clear exactly how to relate performance under noncontingent reinforcement to a judgment of contingency made after the performance, the experiments are suggestive. Wright (1962) used noncontingent schedules of reward in the setting of a trial-and-error problem. With higher frequencies of reward, response patterns were more orderly than they were at intermediate frequencies of reward. Bruner and Revusky (1961) provided the subjects with several telegraph keys. Pressing one of the keys resulted in reinforcement at certain times, but the other keys were nonfunctional. The nonfunctional keys, however, were pressed in systematic patterns during the intervals between reinforcements. When questioned, the subjects reported their belief that the entire response pattern was required to produce the reward. In an unpublished experiment by the senior author, noncontingent reward was used in the setting of a concept formation task. The subjects were shown two-digit numbers and asked to respond with a third number. The experimenter pronounced their answer "correct" or "incorrect" on each trial according to noncontingent random schedules. The subjects formed rules which typically entailed the use of several different arithmetical operations for different types of digit pairs. There were indications that rules were held with greater confidence when the fraction of trials "correct" was higher. Hake and Hyman (1953) had subjects observe a random series of binary digits and try to predict each succeeding digit. They concluded that the subjects responded

as though the series were composed of small subsequences some of which are dependable cues to the future behavior of the series.

If the subjects had been asked about contingency, they might have said that what is about to appear depends upon what has just appeared, although, of course, there is no such dependency in a random series.

The general impression which is conveyed by the results of learning experiments with noncontingent outcomes is that the subjects are surprisingly insensitive to the distinction between contingent and noncontingent arrangements. They tend to behave as though outcomes depend on responses, or as though one symbol can be predicted from another, when the events are in fact independent. Further, it is possible to read into some of these experiments the notion that higher frequencies of reward, or of correct prediction, encourage a belief in contingency.

Although previous work provides little basis for a prediction of how well the contingency between responses and outcomes might be evaluated by the subjects in the present experiment, it does suggest some factors which would be expected to produce distortions of judgment.

It appears that confirming cases are given considerable weight in the judgment, while the role of nonconfirming cases is less clear. However, even if nonconfirming as well as confirming cases were taken into account, it is not difficult to see that under certain conditions, a subject might respond in a way which would generate an excess of confirming over nonconfirming cases even though responses and outcomes were, in fact, entirely unrelated. Suppose that one of the outcomes is preferred over the others, e.g., it represents a score point, and that it is programmed to appear frequently and independently of responses. The response choices made by the subject at the outset will thus be accompanied by frequent scoring. If scores reinforce, the response chosen at the outset is likely to be maintained to the virtual exclusion of other alternatives. The predominance of one response (or pattern of responses) together with one outcome will yield an excess of confirming over nonconfirming cases which, in turn, might lead to a spurious belief in control. The situation is analogous to the previous example in which the predominance of

blue eyes and blonde hair gave an excess of confirming cases even though eye and hair color were independent. The confirming cases are, of course, concentrated in a single cell of a contingency table and the ΔP index remains at 0.

On the other hand, an excess of confirming cases cannot arise in the noncontingent case, no matter how strongly a particular outcome predominates, if response alternatives are used with equal frequencies. Therefore, a change in the character of the subject's task which would lead him to a more balanced use of response alternatives should reduce the presumed tendency for his belief in control to increase with an increasing predominance of one outcome. The discrimination between contingent and noncontingent cases should improve.

This conjecture was examined in the first experiment. The same set of problems was judged by subjects whose objectives while working the problems were set by one of two contrasting instructions: the instruction to score, or to control. Under the score instruction, one of the two outcomes constituted a "score," the other a "no score," and the subject was instructed, in part, to score as often as possible. Under the control instruction, the two outcomes were neutral symbols; and the subject was instructed to learn how to produce each of them at will on any trial. The control instruction was expected to produce a more balanced use of response alternatives and, as a result, more valid judgments of contingency.

An ancillary purpose of the first experiment was to find out if active involvement as a performer in a learning task adversely affects the validity of judgment of control. Paired with each subject who made the responses (active subject) was one who simply watched a display of the responses and outcomes (spectator). Both subjects judged control at the end of the series of trials.

EXPERIMENT I

Method

Subjects. The subjects were 50 college graduates, males and females, employed at the White Plains, New York, office of the Long Lines Division, American Telephone and Telegraph Company. Their ages ranged from 21 to 58 years with a median of 38.

Apparatus. The active subject was seated in front of a control box and a display panel. For subjects under the score instruction, two buttons labeled " R_1 " and " R_2 ", a button labeled "clear", and one labeled "test" were available on the control box. On each trial the active subject made a single response choice, pressing either R_1 or R_2 . The choice was registered immediately on the display panel by the illumination of the numeral 1 or 2. The indication remained until the end of the trial. If for any reason the subject wished to change his response choice at this point, he could do so by pressing the clear button and making a new response choice. He then pressed the test button which was followed immediately by either the "score" outcome (O_1) or the "no score" outcome (O_2) on the display panel. These outcomes were indicated by the illumination of the words score and no score on the display panel. They remained on for 2 seconds, at the end of which all display lights went off, and the apparatus was automatically set for the next trial.

The apparatus for the subjects in the control instruction involved the following modifications. Neutral symbols were used to represent the outcomes: O_1 was shown by a lighted circle, and O_2 was shown by a lighted square. Two additional buttons, referred to as "call" buttons, were made available on the control box. The call buttons were labeled with a square or a circle to correspond to the outcomes. Under the control instruction, the subject indicated, by pressing one of the call buttons at the beginning of each trial, which outcome he was trying to produce on that trial. The called-for outcome was registered on the display by means of small pilot lamps located next to the unilluminated outcome figures. The subject then made a response choice and operated the test button. Thus, under the control instruction, two choices were made on each trial: first a choice of outcome, made by pressing a call button, and then a response choice.

The subject in the spectator position was visually isolated from the active subject, but viewed a duplicate display.

The events displayed were automatically controlled by the subject's responses through relays and a programming device. Operation of the test button activated a teletype reader which read punched paper tape to produce the appropriate outcome. Two outcome sequences were punched on different channels of the tape. The response choice determined which channel was to produce the outcome for that trial. In the case of problems in which outcomes were not contingent upon responses, identical outcome sequences were punched on both channels.

Counters recorded the events of each trial so that the frequency of all response-outcome combinations and, for the subjects under the control instruction, call-response-outcome combinations, could be obtained readily.

Instructions. The instructions were not read to the subject, but they were explained according to a plan to which the experimenter adhered closely.

The same wording was used to express the key ideas to all subjects. Questions were answered as they arose. Spectator subjects listened while instructions were given to the active subjects.

For the score instruction, the task was explained as one of "finding a way to respond which will make the score light appear as often as possible," and for the control instruction as one of "finding a way to control which of the outcomes (square or circle) will appear on any trial." The subjects were then told that at the end of each of five separate problems they were to make a judgment of the degree of control which had been exerted over the outcomes by response choices. They were shown a scale marked at intervals of 10 with extreme values of 0 and 100. The extremes were labeled *No Control* and *Complete Control*. The subjects were then told:

After each problem you are to indicate your judgment of control by putting an "X" some place on the scale: at 100 if complete control has been achieved, at 0 if no control has been achieved, and somewhere between these extremes if some but not complete control has been achieved over the outcomes. Complete control means that you can produce the score light or the no score light (alternatively, the circle or the square) on any trial by your choice of responses. No control means that you have found no way to make response choices so as to influence the outcomes. Intermediate degrees of control mean that your choice of responses influences which outcome appears even though it does not completely determine the outcome.

It should be noted that in instructing the subjects who were in the score condition, it was explicitly stated that *control* means the ability to produce the "no score" light as well as to produce the "score" light. Similarly, in the control instruction it was stated explicitly that control means the ability to produce each of the two outcomes, at will, on any trial.

The subjects were told that any one of the following three states of affairs might be found on any problem: (a) response choices do not affect the outcomes, i.e., there is no control; (b) one response produces one outcome more often than does the other response; or that (c) different patterns of responses produce different outcomes. The possibility that the correct judgment for a given problem might be one of zero control was stated explicitly.

Both spectator subjects and active subjects were offered the option of keeping a record of events. Blank space was available in the test booklet for this purpose. The subjects were told not to look back at earlier records once a new problem had begun.

Problems. A problem consisted of 60 self-paced trials. The statistical structure of the problems is shown in Table 1. Each subject worked five problems. Three were noncontingent (A, B, C), and two were contingent (X, Y). (The pretraining problems shown in Table 1 were used only in Experi-

ment III). Note that the noncontingent problems differ in the degree of bias in the outcome probabilities. Since O_1 stands for "score" in the score instruction, the number of scores in the 60 trials for Problems A, B, C will be, in order, 30, 48, and 8 scores. In the case of contingent problems (X, Y) the number of scores will depend on response choices.

Design and Procedure. The assignment of subjects to the score or control instructions and to the active or spectator positions was made at random. The order of problems was governed by 5×5 Latin squares. Twelve different randomizations of the trial sequence were used on each problem. These randomizations were subject to the restriction that a given outcome had the same programmed frequency in the first and second halves of the 60 trials. In the case of contingent problems, the assignment of conditional outcome probabilities to R_1 or R_2 was interchanged so that each problem was run equally often with R_1 or R_2 leading to O_1 most frequently.

The subjects were given a 10-trial practice run to familiarize them with the operation of the equipment. They recorded their judgments for each of the five problems on a separate page of a booklet. The scale described above was printed on the top of each page preceded by the question: "How much control do you [does the other subject] have over the outcomes?"

Data from two pairs of subjects were excluded from the analysis since in each case one of the subjects indicated in the course of the experiment a gross failure to understand instructions. Usable data were obtained on 10 pairs of subjects under the score instruction and on 13 pairs under the control instruction.

Results and Discussion

Effect of Instruction and Involvement. Results on the judgment of control are given in Table 2. Instructions had a strong effect on judgment. The effect was particularly evident in the case of Problem C where, under the score instruction, the median judgment for active subjects was 0, while under the control instruction it was 55.0. Under neither instruction, however, did judgment follow the ΔP index at all closely. In both cases, some noncontingent problems were judged higher than one of the contingent problems. That there was, in fact, no significant relation of judgment to contingency in either group is supported by correlational data based on individual performance. These data are given below in another connection.

The assumption that, for noncontingent

TABLE 1

CONDITIONAL OUTCOME PROBABILITIES AND ΔP
INDEX FOR TEST AND PRETRAINING PROBLEMS

Problem	Conditional outcome probabilities		ΔP
	Pr(O ₁ /R ₁)	Pr(O ₁ /R ₂)	
Test ^a			
Noncontingent			
A	.500	.500	0
B	.800	.800	0
C	.133	.133	0
Contingent			
X	.800	.500	.3
Y	.800	.200	.6
Pretraining ^b			
Noncontingent			
A'	.500	.500	0
C'	.900	.900	0
Contingent			
Y'	.900	.100	.8

^a Experiments I, II, and III.^b Experiment III.

problems, higher frequencies of scoring would produce a greater concentration on one response alternative was not borne out. An analysis of variance of individual response biases (deviations from an equally frequent use of R₁ and R₂) for Problems A, B, C under the score instruction showed no significant effect of the frequency of scores on bias. A similar analysis of response bias for these problems under the control instruction also showed no effect. The possibility remains that the amount of repetition of a particular *sequence* of responses was affected by the frequency of scores, but this cannot be ascertained with the present data.

In any case, the idea that a response bias would produce an excess of confirming over

nonconfirming cases, and thus lead to a high degree of judged control for noncontingent problems, seems incorrect on another ground. The results for Problem A almost rule out the notion that judgment was based on a balance of confirming and nonconfirming cases. In Problem A the two outcomes occur equally often. This means that, except for sampling errors, the number of confirming cases (the conjunction of one outcome with a particular response or pattern of responses plus the conjunction of the other outcome with some other response) will equal the number of nonconfirming cases, no matter how much bias exists in the use of responses. But, in spite of the equality between confirming and nonconfirming cases, the subjects judged, on the average, that responses controlled outcomes to a moderately strong degree in this problem.

The degree of active involvement had no significant effect on the judgment of control. Of the 10 comparisons of mean judgment made by active as against spectator subjects (5 problems \times two instructions) only a single comparison yielded a value of t for which $p < .05$. This occurred in Problem C under the score instruction in which a very small difference in mean judgment was accompanied by a very low standard deviation of judgment.

The rank-order correlation of the median judgment on each problem for active subjects with spectator subjects was 1.0 under the score instruction and .9 under the control instruction. Thus it is clear that some feature of the problems does result in systematic differences in the degree of judged control.

TABLE 2

MEDIAN, MEAN, AND STANDARD DEVIATION OF JUDGED CONTROL BY PROBLEMS (WITH ΔP VALUES)
AND EXPERIMENTAL CONDITIONS

Problem	ΔP	Score instruction						Control instruction					
		Active subject			Spectator subject			Active subject			Spectator subject		
		<i>Med</i>	<i>M</i>	<i>SD</i>	<i>Med</i>	<i>M</i>	<i>SD</i>	<i>Med</i>	<i>M</i>	<i>SD</i>	<i>Med</i>	<i>M</i>	<i>SD</i>
A	0	20.0	19.9	12.6	20.0	18.1	16.8	30.0	33.5	18.4	27.0	24.2	22.4
B	0	72.5	66.6	16.1	80.0	71.0	19.5	50.0	54.2	24.9	40.0	38.6	23.1
C	0	0	.4	1.2	0	5.3	6.9	55.0	52.4	29.7	50.0	51.4	23.2
X	.3	55.0	55.9	20.1	70.0	59.6	27.8	40.0	35.5	22.2	30.0	31.6	21.4
Y	.6	55.0	56.9	16.5	70.0	58.0	25.6	65.0	61.2	25.0	48.0	43.6	30.1

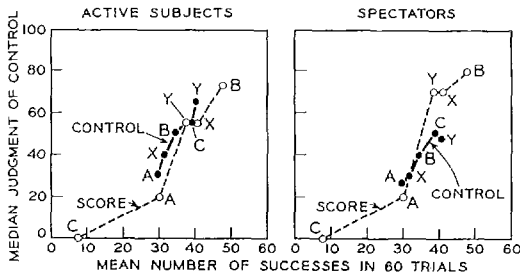


FIG. 1. Median judged control as a function of mean successes.

Prediction of Judgment from Successes. The feature of the problem which best predicts judgment turns out to be the number of successful trials. In the case of the score instruction the number of successes is simply the number of times the score light appears. Under the control instruction, it will be recalled, the subject indicates by means of call buttons the outcome which he is trying to produce on each trial. We may count as a success any trial on which the outcome he is trying for appears.

The relation of median judged control to the mean number of successes is shown in Figure 1. The amount of judged control shows a similar increasing trend against successes under both control and score instructions and for both active and spectator subjects.

The product-moment correlation of individual judgments with number of successes, based on all subjects and all problems, was .70. The correlation was .72 for active subjects and .68 for spectator subjects. A study of the scatter plots for other subgroups, and for contingent and noncontingent problems separately, did not suggest any systematic differences in the regression of judgment on success.

A correlational analysis was also carried out on the relation of judgment to contingency. The response-outcome contingencies in the 2×2 tables which result from the subject's choices and outcomes will differ from the nominal contingencies because of sampling errors. It is therefore possible that judgment is correlated to some extent with the actual contingency even though it bears no relation to the nominal contingency. However, the partial correlations between

judgments and the chi-square values based on actual response-outcome frequencies in the 2×2 tables with successes held constant averaged only .13 for the different groups of the experiment and were in no case significantly different from zero. A similar analysis on the call-outcome contingencies for the control instruction gave an average correlation of only .08.

Of the 46 subjects, 23 made some record of events during the problems. The correlation of judgment with success for all subjects making records, taken over all problems and conditions, was .73, while that for subjects not keeping records was .62.

In summary, the correlational analysis shows no evidence that judgment was systematically influenced by any feature of the problems other than the number of successes in 60 trials, nor by any of the experimental conditions except insofar as these conditions affected the number of successes.

Factors Affecting Frequency of Success. In the case of the noncontingent problems (A, B, C) under the score instruction, the number of scores, and hence the number of successes, is completely determined in advance by the tape program. However, under the control instruction the number of successes (agreements between the called-for outcome and the actual outcome) depends jointly upon the relative frequency with which each outcome appears and with which it is called for. The expected number of successes for a 60-trial problem is given by: $60 [P(C_1) P(O_1) + P(C_2) P(O_2)]$ where $P(C_1)$ is the probability of calling for O_1 and $P(O_1)$ is the probability of O_1 . For any problem with unequal outcome probabilities, the expected number of successes increases as the probability of calling for the more frequent outcome increases.

The subjects did tend to bias calling frequencies toward outcome frequencies as shown by the results given in Table 3. Dunnett's procedure for comparing several means with a control mean (Steel & Torrie, 1960) was used to test the significance of the difference in calling frequencies. The value of 32.4 obtained in Problem A, in which the outcome frequency was unbiased, was the control mean. The departure from this value was significant in Problem C (p

< .05), but not in Problem B. As a consequence of the calling bias, the obtained mean number of successes in Problem C was 39.0 which is significantly larger than the expected number of 30 based on equally frequent calls for O_1 and O_2 ($t = 3.97$, $p < .001$).

By trying more often to produce that outcome which is preprogramed to appear more often, the subjects produce more successes and thus judge a higher degree of control. The tendency to match the probability of predicting an outcome to its probability of appearing is a well-known result when the subject's task is one of prediction (e.g., Grant, Hake, & Hornseth, 1951). The presence of a similar trend when the subjects are instructed to control outcomes suggests that they may fail to distinguish the prediction from the control of outcomes.

In the case of contingent problems under the score instruction, successes increase with the proportion of trials on which the response associated with the higher conditional score probability is used. A significant preference for the response choice associated with the higher conditional score probability did occur for Problem Y (obtained mean frequency of 44.8 against an expectation of 30; $t = 7.57$, $p < .001$) where the contingency is strongest, but not for Problem X (obtained mean frequency of 35.0). Thus, in Problem Y the mean number of successes was increased from the expectation of 30 to an obtained value of 38.6 ($t = 6.83$, $p < .001$).

Results for the control instruction were similar. In Problem Y, those call-response combinations which maximize the expected number of agreements between call and outcome were used with significantly greater frequency than expected by chance (obtained mean of 47.5 against an expectation of 30; $t = 6.16$, $p < .001$). As a result, the number of successes was increased from an expectation of 30 to an obtained mean of 40.3 ($t = 6.83$, $p < .001$).

The results on response choices for contingent problems show that the stronger contingency in Problem Y did have an effect on performance even though, as previously shown, there is no evidence that contingency had any direct effect on judgment.

TABLE 3
MEAN CALL AND OUTCOME FREQUENCIES FOR
NONCONTINGENT PROBLEMS, CONTROL
INSTRUCTION

Problem	Call O_1	Frequency of O_1
A	32.4	30
B	36.0	48
C	16.8	8

Meaning of Judgment. The absence of a relation between judged control and actual contingency in any of the experimental groups makes it quite unclear as to what the subject means by the judgment. One would like to know how the judgment is related to other statements which the subject might be ready to make about the connection between his performance and the outcomes. In particular, does the judgment of a high degree of control carry with it the implication that the proportion of outcomes of a given kind can be greatly altered through responses? Perhaps the subjects take the word "control" to be synonymous with "getting what you want," and not with the ability to alter what you get. If so, the subject might actually have a correct appreciation of the degree to which outcomes can be altered even though his judgments of control are unrelated to contingency. On the other hand, it may be that when the subject judges a high degree of control he also believes that he is able to alter outcomes.

These questions were examined in Experiment II in which the subject was asked both to estimate his ability to alter outcomes and, as before, to judge control.

EXPERIMENT II

Two different sets of questions were used with separate groups of subjects in an attempt to assess the subject's belief in his ability to alter outcomes. The questions were answered at the end of each problem. In one set, referred to as the "switched-intention" set, the subject is first asked to estimate how often he would be able to produce O_1 if he were given another 60 trials. He is next asked how often, given still another 60 trials, he could produce O_2 if he switched his intention to the production of

that outcome. These two estimates can be used to arrive at a subjective ΔP index, or $\Delta P'$. The first answer is taken as an estimate of the probability of O_1 given whatever response or pattern of responses the subject believes is most likely to produce O_1 . The second answer yields, by subtraction from 60, an estimate of the probability of O_1 when the subject is trying to avoid it, i.e., when he is trying to produce O_2 . The value of $\Delta P'$ is the difference between these two conditional probabilities of O_1 . The logic behind the computation of $\Delta P'$, the subjective index, is no different than that behind the computation of the ΔP index of contingency.

Another set of questions is referred to as the "random-player" set. The subject is first asked how often in 60 trials he can produce whatever outcome he feels best able to produce. He is then asked how often that outcome would occur if the response choice had been made by chance, i.e., by the flip of a coin. The difference between the subject's estimate of how often he can produce the chosen outcome, and his estimate of how often it would be produced by a random player, also provides an index of his belief in his ability to alter outcomes.

The actual values of $\Delta P'$ based on the answers to the switched-intention questions might be expected to run higher than the values based on the answers to the random-player questions. Presumably, a greater difference in outcome probabilities is produced by exerting control in two opposing directions (i.e., in the attempt to first maximize one outcome and then to maximize its exclusive alternative) than by exerting control in only one direction and comparing the results against chance.

Method

Subjects. Thirty-two undergraduates at Duke University, males and females, served as subjects.

Procedure. The same apparatus and set of five test problems were used as in Experiment I. All subjects were run individually in the active position, and all made a judgment of control on a scale as in Experiment I. Half of the subjects ran under the score instruction and half under the control instruction.

Under each instructional condition, half of the subjects answered the random-player questions,

and half of them answered the switched-intention questions.

Instructions. Instructions were similar to those in Experiment I with the following differences. No statement was made concerning the possibility that the outcome which appears on a trial depends on response choices for preceding trials. For the control instruction, greater emphasis was placed on the fact that the call buttons were only indicators of intention and had no effect on outcomes. The subjects were required to state correctly the definitions of complete control, no control, and partial control prior to beginning the first problem.

The additional questions were explained to the subject in advance as was the scale for the judgment of control.

Design. Eight subjects were run in each of the four experimental conditions (score and control instructions each with switched-intention and random-player questions). The subjects were assigned to these conditions at random. Within each group every subject received the five problems in a different order, creating an approximate balance in the frequencies with which problems appeared in each ordinal position. Eight different random tapes were used for each problem. The same set of orders and randomizations was used in each of the four groups.

Results and Discussion

Relation of $\Delta P'$ to Judged Control. No significant correlation between judgment on the scale of control and $\Delta P'$ values was found within any of the four experimental groups. Indeed, the scale values for judged control associated with $\Delta P'$ values of zero were scattered over the range of the scale from zero control to almost complete control.

The scale judgments compared quite closely with those obtained in Experiment I, and appeared to follow the same increasing trend with number of successes. The $\Delta P'$ values showed no relation to successes nor to actual contingencies. For example, the number of subjects who gave higher mean values for $\Delta P'$ on contingent than on noncontingent problems was no greater than expected by chance (18 out of 32 was obtained against the chance expectation of 16).

The $\Delta P'$ values were extremely erratic. Whereas a representative value for the coefficient of variation based on judgment of control would be 50%, a typical value for $\Delta P'$ would be in excess of 100%. Further,

whereas Kendall's index of concordance, W , based on the rank orders given to problems by different subjects, showed significant concordance in all four experimental groups when the ranks were based on judged control ($p < .01$ in each case) it failed to reach significance in any group when the ranks were based on $\Delta P'$. Thus, we are unable to distinguish with confidence the rankings for problems based on $\Delta P'$ values from random assignments of ranks to problems. Contrary to expectation, the overall mean value of $\Delta P'$ based on the switched-intention question was not significantly higher than that based on the random-player questions.

The lack of correlation between judged control and $\Delta P'$ values leads to the conclusion that the subject's concept of control is not typically equivalent to "the ability to alter outcomes." Further, the lack of relation between $\Delta P'$ and actual contingency indicates that the subjects do not have a correct appreciation of their ability to alter outcomes.

Problem of Instructions. The possibility remains that the subjects do have a concept similar to that of control in the sense of contingency, but that the instructions have failed to bring that concept into play.

The question of how to instruct is a particularly difficult one in the present context. It would, of course, be possible to instruct the subject in an explicit procedure for calculating or estimating $\Delta P'$, but this would tell us very little about his everyday concept of control. The approach taken in Experiment III to the problem of how to give a clearer instruction without providing an explicit rule of calculation was to give the subject prior experience with an example of zero control and of strong control. If the subject does have a concept of control in the sense of contingency, perhaps it can be brought into play by this small amount of pretraining.

EXPERIMENT III

It would appear from the results of the two previous experiments that the examples used for the purpose of instruction must, if they are to lead to valid judgment, counter-

act the tendency to judge control in terms of successes. Further, if the $\Delta P'$ values are to become consistent with judged control, it will be necessary to jointly specify correct values for judged control on the scale, and to correct answers to the additional questions from which the $\Delta P'$ values are obtained.

This line of reasoning led to the use of three different types of pretraining. One group of subjects received examples which were chosen so that the number of successes that would be achieved was correlated with the correct values for the amount of control. In this group, judged control on test problems should show, as in the previous experiments, a dependence on successes since the pretraining does nothing to undo the relation of judgment to success. In a second group, the pretraining examples were chosen so that the number of successes would not vary with the correct values for judged control. This might lead to valid judgment since the tendency to judge on the basis of success should be counteracted. Finally, a third group received the same pretraining examples as did the second group, but in addition, was given correct answers to the questions from which the $\Delta P'$ values are computed. Valid judgments of control are, presumably, most likely in this group, since correct estimates of the alterability of outcomes are given together with the correct values for judged control. The equivalence of control and alterability should be emphasized by this procedure.

All subjects were given the control instruction. They worked two pretraining problems. In one of these there was no contingency, while in the other there was a rather strong contingency. The correct judgments for these problems were shown to the subject in advance. The presence or absence of a correlation of success with control was manipulated by variations in the pretraining problem which exemplified no control. It was known from Experiment I that under the instruction to control outcomes, the subjects tend to match the frequency of calling a given outcome to the frequency with which that outcome appears. As a consequence, the number of suc-

TABLE 4
MEAN SUCCESSES ON PRETRAINING PROBLEMS

Group	Problem		
	A'	C'	Y'
I	28.4	—	42.8
II	—	41.4	41.2
III	—	42.1	45.4

cesses increases with an increasing predominance of one outcome. It was also known that the subjects take advantage of a moderately strong contingency of outcomes upon responses to increase the number of successes. Therefore, it should be possible to produce about the same number of successes in a noncontingent problem with disproportionate outcome frequencies as in a moderately contingent problem. The pretraining problems for Groups II and III in the present experiment were selected to produce this result, and thus to remove any correlation between contingency and the mean number of successes.

On the other hand, when the noncontingent pretraining problem has equally frequent outcomes, the mean number of successes will be less than for the contingent problem. Contingency and the mean number of successes will, in this case, vary together. The pretraining examples in Group I were selected to achieve this result.

Method

Subjects. Twenty-four undergraduates at Duke University, males and females, served as subjects.

Procedure. The apparatus was the same as in previous experiments. The same five test problems, problem orders, and randomizations were used as in Experiment II.

The pretraining problems each consisted of 60 trials. Their statistical structure is shown in Table 1. All three groups received, as an example of moderately strong control, Problem Y' which is similar to Problem Y of the test series, but represents a somewhat stronger contingency. It should, on the basis of previous results, yield about 40 successes in 60 trials. The example of zero control for Group I was Problem A'. It is identical with Problem A of the test series, it has equal outcome frequencies, and it will yield an average of about 30 successes in 60 trials. In Groups II and III the example of zero control was Problem C' which is similar to Problem C of the test series, but it has slightly more disproportionate outcome frequencies. Results of the previous experiments

predict an average of about 40 successes for Problem C'.

Instructions. All subjects were run under the modified control instruction used in Experiment II. They all answered the random-player questions in addition to making a judgment of control by marking the scale. Prior to working pretraining problem Y', the subject was told: "You will have very good control over the outcomes by your choice of responses." He was then shown a sample answer sheet with the scale of judged control marked at 80. Prior to working the pretraining problem exemplifying zero control the subject was told: "Your choice of responses will have no influence over which outcome will appear." He was then shown a sample sheet with the scale of judged control marked at zero.

Group III received the following additional information concerning correct answers to the random-player questions. For Problem Y', the subject was told:

If you were to try for the circle on each of 60 trials you could make it appear 54 times. A coin-flipping player would get the circle only 30 times. If you had decided instead to go for the square, you could make that outcome appear 54 times out of the 60 trials. And, of course, the coin-flipper would get just 30 squares.

For Problem C', the subject was told:

No matter which outcome you try for, the square will appear 54 times in the 60 trials. The random player would also get 54 squares in the 60 trials.

Design. Eight subjects were run in each group. Four different randomizations of each pretraining problem were used. For Problem C', O₁ appeared most frequently for half the subjects while O₂ appeared most frequently for the remaining subjects. For Problem Y', R₁ led to O₁ most frequently for half of the subjects, while R₂ led to O₁ most frequently for the remaining subjects.

Results and Discussion

Successes on Pretraining Problems. The mean number of successes on pretraining problems is shown in Table 4. The results were as anticipated. The mean number of successes on Problem A' was well below the value for Y', producing the desired correlation between success and control in Group I. The mean number of successes on Problem C', on the other hand, was very close to the value for Y'. Therefore, as was intended, successes did not vary systematically with control in Groups II and III.

Judgment of Test Problems. The discussion of results centers on the effects of the experimental conditions on the correlations between the variables which appear in the column headings of Table 5.

The mean correlation of judgment with success was significantly higher in Group I than was the overall mean correlation for Groups II and III combined ($p < .025$, one-tailed U test based on individual rho's). Thus we obtained the anticipated reduction in the correlation of judgment with success as the result of identifying a problem which yields frequent success as one having zero control. Ranks based on the judgment of control showed significant concordance in Group I ($p < .05$), but not in Groups II or III.

The correlation of $\Delta P'$ with success was moderately high in Groups I and II. It was lower and not significantly greater than zero in Group III in which the correct $\Delta P'$ values were specified in pretraining. The rankings based on $\Delta P'$ were not significantly concordant in any group.

The mean correlation of judgment with $\Delta P'$ was substantial in Groups I and III and significantly lower than for either of these in Group II ($p < .025$ by one-tailed U test). An interpretation of this pattern is that the correlation is high in Group I largely because both judgment and $\Delta P'$ are following success: a result also obtained in the comparable group in Experiment II. The correlation is high in Group III because of the joint specification of correct values for $\Delta P'$ and for judged control in pretraining. In the absence of these special circumstances $\Delta P'$ and judged control are not in agreement.

The correlation of judged control and of $\Delta P'$ with the ΔP index of contingency is given in the last two columns of Table 5. These correlations are low, and in all cases they are not significantly different from zero. It is apparent that the pretraining conditions for Groups II and III yield no improvement in the validity of judgment or of $\Delta P'$ values.

The results of Experiment III go against the notion that the failure to find valid judgments in the first two experiments was due to a lack of communication. Even with appropriate pretraining, no significant correlation appears between the ΔP index of actual contingency and judged control, or between the ΔP index and the subject's es-

TABLE 5
MEANS OF INDIVIDUAL RHO'S

Group	Variables Correlated				
	J-S	$\Delta P'$ -S	J- $\Delta P'$	J- ΔP	$\Delta P'$ - ΔP
I	.77*	.66*	.91*	.25	.01
II	.18	.48*	.15	.01	.16
III	.38	.24	.75*	.25	.02

Note.—Abbreviations are: J—judgment, and S—success.

* Significantly different from 0 at .05 level by binomial test for number of individual rho's > 0 .

timates of his ability to do better than a random player.

The conditions of pretraining did, however, have an effect. When the pretraining problems were selected to produce a covariation of successes with actual control (Group I), judged control increased with successes, a replication of the results of the first two experiments. When, however, the pretraining problems produced the same mean level of success on the example of no control as on the example of strong control, the tendency to judge control on the basis of successes was removed.

The results for the $\Delta P'$ values are similar to those of Experiment II. These values again fail to show significant concordance; and in Group II, they also fail to correlate significantly with judged control. Thus it is again found that the subject's judgment of control does not predict his estimate of the degree to which he can alter outcomes.

QUESTIONNAIRE RESULTS: EXPERIMENTS I, II, and III

In each experiment subjects answered a written postexperimental questionnaire which elicited certain background data and information concerning the basis of judgment.

Of the total of 102 subjects in all experiments, 19 reported having taken at least one college-level course in statistics. Seventeen of these were in Experiment I. The pattern of judged control for these subjects was similar to that given by subjects without statistical training.

All postexperimental questionnaires contained the following item:

TABLE 6

ΔP INDEX, SUCCESSES, MEDIAN JUDGMENT OF CONTROL, AND NUMBER OF SUBJECTS ELECTING TO JUDGE FOR HYPOTHETICAL PROBLEMS IN THE POSTEXPERIMENTAL QUESTIONNAIRE, EXPERIMENT II

Problem	ΔP	Score instruction			Control instruction		
		Suc- cesses	Mdn. Judg- ment	N ^a	Suc- cesses	Mdn. Judg- ment	N ^a
1	0	48	77	15	39	50	12
2	.5	35	35	13	32	50	12
3	.7	40	50	12	18	30	13

^a N of subjects electing to judge.

If you were to observe someone else working a problem in which on every trial he made the same response and got the same outcome, how much control would you say he had over the outcomes?

The subjects checked the following alternatives with the frequencies given in parentheses: *complete control* (49), *medium control* (15), *no control* (8), *uncertain* (30).

In Experiment I, in which the possibility of sequential dependencies was stated explicitly, 40 of the 46 subjects reported their belief that outcomes on a trial depended in various ways on the events of preceding trials. In Experiments II and III, no statement was made to the subject about sequential dependencies. In Experiment II, 16 of the 32 subjects indicated their belief in such a dependency; while in Experiment III, 10 of the 24 subjects did so. Inspection of the judgments made by those who reported sequential dependencies and by those who did not, revealed no systematic differences.

The questionnaire for Experiment II included hypothetical sets of data for three 60-trial problems. For subjects under the score instruction these data were the four cell frequencies of the 2×2 table. For those under the control instruction, frequencies were given for the eight possible call-response-outcome combinations. The subjects were asked to make a judgment of control for these data. However, they were given the option of omitting judgment on any problem for which they had no idea how to proceed. The results are shown in Table 6. The influence of success on judg-

ment of control and the failure of judgment to parallel the ΔP index of contingency is evident from these results.

The appearance of a tendency to judge control on the basis of success on the questionnaire is interesting. It parallels the results for the test problems and thus suggests that these results may be quite general rather than being a consequence of certain special features of the present experimental task. The subject is not required to retain serially acquired information in answering the questionnaire, nor does he have any role in producing the data to be judged. Perhaps most significant is that the relation of judgment to success appears even when the subject is provided with a tabulation of event frequencies in the appropriate categories. When the judgment is made on the basis of a tabular summary of frequencies rather than after experience with a series of events, there is no opportunity for erroneous beliefs about the efficacy of *patterns* of choices to enter into the judgment. Since the relation between successes and judged control continues, it would appear that the belief in response patterns is not a necessary condition of the correlation of judgment with success.

CONCLUDING DISCUSSION

The main finding of these experiments is that the amount of judged control was a function of the frequency of successful outcomes rather than of the actual dependency of outcomes upon responses. The relation of successes to judgment is robust since it appears when the subjects work with neutral outcomes (control instruction) as well as with favorable and unfavorable outcomes (score instruction). It appears when the relevant events are simply observed, as well as when they are produced by responses. Further, the subjects who kept trial-by-trial records were no less subject to the effect than those who relied on their unaided memory. Finally, successes continued to have their effect when the judgments were made from an appropriate summary tabulation of the events rather than from an unprocessed trial-by-trial sequence.

The fact that the subjects mark a scale

in response to a question about control does not mean that they have a concept of control which entails the core concept of contingency. It now seems unlikely that the typical subject in these experiments has such a concept. Not only is a high degree of control often judged in the absence of contingency, but the judgment, however arrived at, does not consistently imply for these subjects the ability to alter outcomes through response choices. This conclusion is based on the failure of judged control to stand in any sensible relation to the subject's estimate of his ability to outperform a random player or to change at will the proportion of outcomes of one kind. While it may be that the concept of contingency could have been evoked by other questions and instructions, the failure of pretraining in Experiment III to do so argues that a simple lack of communication was not responsible.

The conclusion that the subjects in the present experiment were without a concept of contingency is not intended to preclude the possibility that far more valid judgments of the same statistical structures could be made if the events were cast in a different context. An example of such a context might be one in which inputs were represented as the presence or absence of a drug and outputs as recovery or nonrecovery from infection. A conclusion that is, however, warranted by the results of this experiment is that the typical subject in this population did not have an *abstract* appreciation of statistical contingency. As has been noted, Smedslund (1963) stated a similar conclusion:

normal adults with no training in statistics do not have a cognitive structure isomorphic with the concept of correlation.

It might be added from the results of the present experiment that training in statistics will often fail to improve matters.

We are left, however, with the finding of Inhelder and Piaget to the contrary, namely, that correlational reasoning often appears by the age of 14 or 15 years. It has been seen, however, that the formulation of the concept of correlation by these authors is correct only for a special case

and that it is doubtful that their subjects were using the more general concept of correlation. Further, the context and the method of obtaining the judgment in the present experiments were quite different from those in Inhelder and Piaget. Differences of this kind may have strong effects upon the level of reasoning in the judgment of relationship.

If one were to generalize the present results broadly, one would be left with the puzzle of how people get along as well as they do even though they are unable to judge correctly that some event is controlled by or, on the other hand, is independent of some other event. Surely the distinction has implications for adaptive behavior. The puzzle may be lessened by considering some of the ways in which the present experimental task is not representative of the natural conditions of such judgments. Two features, which may be particularly important, are the absence of relevant temporal variations between input and output, and the discrete nature of the binary input and output.

The temporal succession of input and output was the same for all degrees of contingency and was thus irrelevant. Under natural conditions, however, temporal proximity is undoubtedly an important determiner of the judgment or perception that events are related. It appears to be generally the case that events which are in a statistical sense highly contingent upon some antecedent, also tend to follow that antecedent closely in time.

In the present task, the input events R_1 and R_2 were discrete as were the outputs O_1 and O_2 . As a consequence, three states were actually involved in both the input and output since there was also the between-trial state in which neither of the alternate inputs and neither of the outputs occurred. This is to be contrasted with the case in which the input and output are of the on-off type, so that a momentary input appears against a background of nonoccurrence and is followed, with some probability, by an output event which also appears as the interruption of a resting state. Perhaps such a context is more representative and would lead to more valid judgment. It

is also in this context that temporal variables would normally become important.

The features of the present task which seem to militate against valid judgments are related to the rough distinction made at the outset of this report between the perception and the judgment of a relation. It was suggested that the term perception applies to the case in which the awareness of a relation follows immediately upon the joint occurrence of two events which rarely appear alone, whereas the term judgment applies when antecedent and consequent often occur alone and a series of observations is necessary in order to estimate whether the frequency of joint occurrence exceeds the chance expectation of joint occurrence. The distinction is, of course, related to degree of contingency, since the conditions under which one may speak of the perception of a relation are generally those associated with strongly contingent events. One could still speak in such cases of a comparison of conditional probabilities, i.e., of the probability of the event with, as against without, the antecedent. However, the appreciation of the probability of the event in the absence of the antecedent would be based on prior experience of long standing and would be more like an expectation than an estimate. Thus, from a psychological point of view, it may be misleading and artificial to view the perception of contingency in terms of a comparison of probability estimates.

The present results do not in fact suggest that comparisons of probabilities played much, if any, role in mediating the judgment. Rather, it appears that instead of making comparisons between events within the task, control was judged in terms of the degree of success in the performance of the task as a whole. It was, for example, not uncommon for subjects to speak of having control over just one of the outcomes; a remark which is incompatible with a judgment based on differential effects of responses within the task itself. It is as though the subjects were evaluating their performance against some expectation of how often a favorable event would occur if responses had no control over outcomes.

Many subjects were apparently judging control against a base-line expectancy of zero successes. Even with only 8 successes in 60 trials (Problem C under score instruction) almost half of the subjects judged some nonzero level of control.

An expectation of zero successes in the absence of control could be understood as a generalization from common experience. In ordinary commerce with the environment the joint occurrence of some action and a favorable event (or, more broadly, an event upon which attention has been focused) almost always represents a contingent or causal relation. Chairs do not often move unless pushed, lights do not often come on until the switch is thrown, and so on. In these cases the assumption that the event never occurs until caused is generally correct. Control over a single outcome is perceived against a resting state of no occurrence. When the assumption of a zero base line is altogether inappropriate, such as in games of chance, casual observation as well as the present results suggest that erroneous beliefs in controlling or contingent relations are prevalent.

SUMMARY

Three experiments were conducted to examine subjects' beliefs in the degree of control exerted over outcomes through response choices when outcomes were or were not contingent upon responses.

All subjects worked a set of two contingent and three noncontingent problems in a two-response, two-outcome situation. After each problem the subject judged the degree of control exerted by his responses over outcomes and in Experiments II and III also made certain estimates relating to his ability to manipulate outcome frequencies. The subjects were told in advance that they were to judge control and that for some problems the correct judgment might be one of zero control.

In Experiments I and II, judgments were obtained from subjects who made response choices or who were only spectators. Judgments were also obtained under the instruction to produce as many scores as possible (score instruction) or under the instruction

to control the appearance of neutral outcomes (control instruction). In either case, certain events may be defined as successes. Under the score instruction a success is simply the appearance of a "score" light; while under the control instruction it is an agreement between the outcome which the subject is trying to produce on a given trial and the outcome which appears.

The judgment of control was positively correlated with the frequency of success in all conditions. It was not systematically influenced by the presence versus the absence of contingency or by the other experimental variations except insofar as these affected the frequency of success.

In Experiment III the effect of a limited amount of pretraining on judgment was examined. All subjects worked a contingent and a noncontingent sample problem for which correct judgments were specified in advance. The use of pretraining problems in which the frequency of success was greater on the contingent than on the noncontingent sample resulted in a correlation of judged control with success similar to that obtained in the previous experiments. When the structure of the pretraining problems led to approximately the same number of successes on contingent and noncontingent problems, the correlation of judgment with success was significantly re-

duced. The validity of judgment was, however, not improved.

Whereas there was significant concordance among the subjects in their judgment of control (except where pretraining reduced the correlation of judgment with success) a measure of manipulatability derived from the subjects' estimates of their ability to produce given outcomes consistently failed to show a significant degree of concordance. This measure was generally not in agreement with judged control, nor did it accurately reflect the presence versus the absence of contingency.

The consistent failure to discriminate contingent from noncontingent structures and the lack of agreement of formally equivalent measures of the subject's beliefs concerning the control of outcomes by responses suggests that erroneous beliefs concerning control may be traced to the absence of a statistical concept of contingency in untutored subjects. There is suggestive evidence that the subjects do not distinguish the ability to manipulate outcomes from their ability to predict them. The base line against which the subjects assess their performance appears to be one of zero occurrence of the event of interest in the absence of personal causation. This base line, which is inappropriate in the present context, may arise through a generalization based on everyday experience.

REFERENCES

- BRUNER, A., & REVUSKY, S. H. Collateral behavior in humans. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 349-350.
- GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *Journal of Experimental Psychology*, 1951, **42**, 1-5.
- HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *Journal of Experimental Psychology*, 1953, **45**, 64-74.
- INHELDER, B., & PIAGET, J. *The growth of logical thinking from childhood to adolescence*. New York: Basic Books, 1958.
- SMEDSLUND, J. The concept of correlation in adults. *Scandinavian Journal of Psychology*, 1963, **4**, 165-173.
- STEEL, R. G. D., & TORRIE, J. H. *Principles and procedures of statistics*. New York: McGraw-Hill, 1960.
- WRIGHT, J. C. Consistency and complexity of response sequences as a function of schedules of noncontingent reward. *Journal of Experimental Psychology*, 1962, **63**, 601-609.

(Received October 28, 1963)