Memory & Cognition
1980, Vol. 8 (5), 459-467

A rule analysis of judgments of covariation between events

HARRIET SHAKLEE and DIANE TUCKER
University of Iowa, Iowa City, Iowa 52242

Several strategies are proposed as bases for judgments of covariation between events. Covariation problems were structured in such a way that patterns of correct and incorrect judgments would index the judgment rule being used by a given subject. In two experiments, 10th-grade or college subjects viewed a set of covariation problems, each of which consisted of a set of observations in which each of two events was defined as present or absent. Subjects were asked to identify the relationship between the events. Subjects' response patterns suggested that the modal strategy was to compare frequency of confirming and disconfirming events in defining the relationship. Response accuracy was influenced by pretraining on the concept of covariation and by response format. Instructions to sort the observations did not influence judgment accuracy.

Research investigating people's competence at judging correlations between events has resulted in a maze of contradictory results. In the basic paradigm, subjects are presented data instances illustrating one of two event states (e.g., presence or absence) for each of two events. The subject's task is to identify the direction and/or strength of the relationship between the events. Inhelder and Piaget (1958) and Seggie and Endersby (1972) each found accuracy to be the norm among adolescent and adult subjects identifying such relationships. Others (e.g., Jenkins & Ward, 1965; Niemark, 1975; Smedslund, 1963) have found such accuracy to be rare among populations comparable in age and expertise.

Comparisons between studies identify procedural differences that may account for such performance variability. Most notably, some researchers have presented data in summary tables (Seggie & Endersby, 1972; Smedslund, 1963); others have presented the data as a series of individual instances. Ward and Jenkins' (1965) systematic comparison of tabled and sequential data presentations demonstrates better judgment performance in the tabled format. However, performance differences occur even among experiments using similar procedural conditions. Thus, Seggie and Endersby's (1972) and Smedslund's (1963) subjects evidenced widely disparate competence, although in both cases the researchers tested nursing students judging health-related problems under similar test conditions. Similarly, comparable conditions in two of Ward and Jenkins' experiments (Jenkins & Ward, 1965; Ward & Jenkins, 1965) produced different judgment patterns, although in each case subjects made contingency judgments about events in a trial-by-trial presentation format. Although judgment conditions clearly do influence performance level, these results suggest that much performance variability remains to be explained.

Experimenters' suggestions about the bases of judgments are as various as the results for which they are meant to account. Several strategies have been proposed as common approaches to judgments of event contingencies.

Researchers have often cited a tendency to focus on the cooccurrence of the target events (Cell a in a traditionally labeled contingency table) as a bias in such judgment, failing to realize the equivalence of joint event nonoccurrences (contingency table Cell d) in defining the relationship. Inhelder and Piaget (1958) identify this as the strategy used by younger adolescents (12-13 years) in their tasks. Smedslund (1963) suggests that the strategy is typical among adults as well.

Inhelder and Piaget (1958) suggest that, by later adolescence (14-15 years), subjects use information in all four contingency cells in defining the relationship. Specifically, covariation is defined as the difference between the sums of events confirming (Cells a and d) and disconfirming (Cells b and c) the relationship. Relationship direction is determined by the sign of the difference. Inhelder and Piaget identify this approach as characteristic of formal operational thought.

Jenkins and Ward (1965) suggest that Inhelder and Piaget's (1958) definition of appropriate contingency judgment is itself limited. Specifically, the Piagetian formula of the difference between frequencies of confirming and nonconfirming cases serves as an effective index of contingency only when the two states of at least one of the variables occur equally often. Otherwise, a correlation may be indicated when, in fact, independence is the case. Thus, whereas the judgments of Inhelder and Piaget's adolescents may have met their criterion, these judgments may have been the product of a limited concept of correlation. Instead, Jenkins and Ward suggest

Reprint requests should be sent to Harriet Shaklee, Department of Psychology, University of Iowa, Iowa City, Iowa 52242.

  0090-502X/80/0459-09$01.15/0

that contingency is best evaluated by comparing the conditional probabilities of an event, given the alternative states of the second event $[P(A_1/B_1) - P(A_1/B_2)]$. By definition, independence is indicated by equivalence between the conditionals. Nonindependence is indicated by any difference.

Such a variety of outcomes and proposed strategies demonstrates the poor consensus about mature competence at a covariation judgment. Despite such conflict, the same literature suggests a possible resolution. That is, covariation judgment is discussed in terms of three different strategies, each of which may result in better than chance performance. The optimal strategy is that suggested by Jenkins and Ward (1965), that is, comparing the conditional probabilites $[P(A_1/B_1) - P(A_1/B_2)$ (conditional probability strategy)]. Intermediate in utility is the strategy suggested by Inhelder and Piaget (1958) as the formal operational solution, that is, comparing the sums of the diagonal cells of a contingency table $[(a + d) - (b + c)$ (sum-of-diagonal strategy)]. This strategy will correctly identify most contingencies, but it may result in errors when the frequencies of alternative states of Factor A (and/or Factor B) are disparate. The least general in utility is the strategy used by Inhelder and Piaget's younger adolescents, that is, judging according to the frequency in contingency table Cell a. Such a rule user would judge the direction of relationship according to whether Cell a was the largest (or smallest) of the cells (Cell a strategy). As a strategy that attends to some relevant information, it may result in correct evaluation of some relationships, but it would result in error when there is a large difference between frequencies in Cells a and d.

According to this analysis, solution accuracy may depend on the particular correlational problem a subject is judging. Problems can be identified that would be solved accurately by all three strategies, resulting in high solution rates. Alternatively, problems solvable only by the more sophisticated strategies may result in high error rates. In fact, by structuring a set of such problems, we can use a subject's solution pattern across problems to identify the specific strategy he or she is using.

Ward and Jenkins (1965) made an early attempt at a similar strategy classification. They proposed a set of seven strategies that might account for subjects' contingency judgments, identifying the judgment each strategy would produce on a series of contingency problems. For each subject, the investigators determined the correlation between actual judgments and those predicted by each of the seven strategies. An individual was categorized as using the strategy whose predicted judgments correlated most highly with his or her actual judgments. Although the mean correlations between subjects' judgments and best rule were high (.84-.95, depending on the strategy classification), the experimenters did not indicate how these correlations compare with those of the next-best rule classification. This is a particular problem, since intercorrelations between judgments predicted by different rules was often high (most rs > .50; some were as high as .80). Any attempt to discriminate between such highly related rules requires evidence not only that judgments are congruent with one rule, but also that they are incongruent with alternative rules. A second problem with the approach is that several of the proposed rules predict subjects' judgments about relationship strength, but they give no information about the direction of the relationship. Relationship direction is a key component of event contingencies and should be clearly specified by judgment models. In fact, accuracy of direction as well as strength judgments can be used to index the subjects' bases for contingency judgments.

A primary focus of this investigation was to develop a set of problems to reliably discriminate between possible judgment rules. Conditional probability, sum-of-diagonals, and Cell a strategies were proposed as rule candidates. Problem sets were identified that would result in different judgment patterns depending on the strategy used.

A set of such problems is illustrated in Table 1a. The problems are hierarchically structured such that Cell a strategy problems are solvable by all three strategies; sum-of-diagonal problems are correctly solved by sum-of-diagonals and conditional probability strategies, but they result in error by Cell a strategy; conditional probability problems are correctly solved by that strategy alone. Solution accuracy was indexed by the direction of the judged relationship (i.e., A_1 associated with B_1, A_1 associated with B_2, A and B unrelated). A subject's solution pattern on the set of problems indicates the strategy used. Problems on the first row of Table 1 illustrate judgments predicted by each of the proposed rules. Although all problems in the row are noncontingent relationships, a subject using the Cell a strategy should identify only the first relationship as noncontingent (Cell a is neither larger nor smaller than the other cells). A judge using the sum-of-diagonals rule should see the first two problems as noncontingent relationships but should judge the last problem as one in which A_1 tends to be associated with B_1 $[(2 + 9) - (6 + 3)]$. A conditional probability judge should correctly identify all of the first-row problems as noncontingent relationships. Table 1b summarizes the solution pattern congruent with each strategy type. Since problems were structured hierarchically, subjects' solution patterns should conform to a Guttman scale. Subjects who fail to solve problems of any strategy type may be using some other strategy (and a poor one at that) or no consistent strategy at all (Strategy 0).

A second interest of this investigation was in the level of everyday sophistication about the concept of correlation. Previous evidence of subjects' errors in contingency judgment may be the result of a variety of conceptual or information processing problems. Ward

Table 1

(a) Cell Frequencies Used for Each Problem Strategy Type: Experiment 1

Cell a Problems			Sum-of-Diagonal Problems			Conditional Probability Problems		
	B_1	B_2		B_1	B_2		B_1	B_2
A_1	6	6	A_1	4	4	A_1	2	3
A_2	6	6	A_2	8	8	A_2	6	9
	B_1	B_2		B_1	B_2		B_1	B_2
A_1	11	4	A_1	8	1	A_1	3	6
A_2	1	8	A_2	4	11	A_2	5	10
	B_1	B_2		B_1	B_2		B_1	B_2
A_1	1	8	A_1	4	11	A_1	2	7
A_2	11	4	A_2	8	1	A_2	4	14

(b) Strategy Use and Resultant Patterns of Problem Accuracy

Subject Strategy Type	Problem Strategy Type: Cell a	Sum of Diagonals	Conditional Probability
Conditional Probability	+	+	+
Sum of Diagonals	+	+	0
Cell a	+	0	0
Strategy 0	0	0	0

Note—+ = accurate; 0 = inaccurate.

and Jenkins' (1965) experiment indicates that stimulus presentation conditions can influence subjects' judgment sophistication. This experiment will examine other potential sources of judgment bias.

Past records of poor performance could be a result of subjects' disinclination to organize the frequency information into a 2 by 2 contingency table. Smedslund (1963) notes that fewer than 20% of his subjects organized their data cards into a 2 by 2 matrix for subsequent judgments. This failure to sort relevant information spontaneously may result in poor estimates of relative frequencies, which would result in inaccurate judgments even by the best of decision rules. Instructions to sort the data into a 2 by 2 table should ameliorate this potential problem. We investigated the influence of such instructions on subjects' performance in this experiment. A comparison group was not so instructed. Superior performance by instructed subjects would implicate poor organization strategies as a source of past performance errors.

Alternatively, subjects' problems may stem from a true misunderstanding of the concept of covariation. In fact, this is one of the few points of agreement in past research on the subject. Although the proposed biases vary, the dominant account of poor performance is that subjects have a basic conceptual misunderstanding of the nature of covariation (Jenkins & Ward, 1965; Smedslund, 1963; Ward & Jenkins, 1965). If this claim is true, such confusion may be corrected by training subjects about event contingencies.

Although most past research has studied untrained judgments, Jenkins and Ward (1965) did incorporate a training paradigm in an attempt to improve judgment. Subjects in their experiment judged two training problems, with feedback as to the correct solution. Although this experience did reduce use of a common erroneous strategy (percent success), it did not increase appropriate rule use. This partial training success may indicate that a more elaborate training effort would result in increased judgment accuracy. A reliable improvement by trained subjects would implicate conceptual misunderstanding as a source of judgment problems. At the same time, such improvement would indicate that subjects are capable of applying an appropriate decision rule once they know the relevant principles. This investigation trained subjects in an introductory discussion on the concept of covariation. The presentation focused on the variety of possible directions and strengths of relationships between events. A comparison group had no such discussion. The two concept training conditions were crossed with the two data-sort conditions to produce four judgment conditions.

Two experiments were conducted to investigate these questions. Consistent with Piagetian work on correlational judgment, high school students were the subjects in Experiment 1. Experiment 2 replicated and extended the investigation to college-aged subjects.

EXPERIMENT 1

Method

Problems. Nine different problem contents were developed, each of which consisted of a set of observations picturing one of two event states for two potentially related everyday events. Three of the problems pictured cakes that either rose or fell in association with one of two states for oven temperature (200 deg or 350 deg), addition of baking powder (some or none), or amount of flour added (large or small). In three other problems, plants were pictured as healthy or sick as a possible function of type of plant food (white or black), light conditions (placed by window or wall), or amount of water (large or small). In the remaining three problems, people (or animals) were pictured as sick or healthy as a possible function of presence or absence of a shot, liquid medicine, or a pill. Each paired observation of event states on the two variables was pictured on a 5 x 8 in. card.

A problem was constructed by combining the appropriate frequencies of each event-state combination into a randomly ordered deck. Each subject judged nine such covariation problems. Table 1a lists the actual frequencies used for each strategy type. Cell a and sum-of-diagonals problems included one noncontingent and two contingent ($\phi = .53$) relationships for each strategy type. Conditional probability problems were all noncontingent relationships due to the limitation imposed by the constraint that the conditional probability solution be different from solutions by other strategies. (It turns out that we were wrong about this limitation—see Experiment 2.) Direction of relationship (A_1 associated with B_1, B_2, or no relationship) was counterbalanced between subjects for each problem content.

Problem sequence was determined by grouping the nine problems into three problem blocks, including one problem from each of the three strategy types. The three blocks were presented in one of three orders, each of which began with a problem of a different strategy type.

Each problem was introduced by a paragraph describing the two events and possible relationships between them. Subjects were instructed to look through the cards to identify the relationship between the events pictured. An example introduction is given in Table 2.[1] A similar introductory paragraph and response scale were developed for each problem content. In each case, subjects indicated that Event A_1 was associated with Event B_1 or Event B_2 or that there was no relationship between them. The reader may examine actual problem frequencies in Table 1a to note the responses that would be predicted by the proposed strategies. Data cards were available to the subject throughout the problem.

Procedure. Subjects were tested individually. Judgment conditions were defined according to two factors, which were crossed to produce four different judgment conditions.

Concept training vs. no concept training. Subjects in the concept training condition began the session with a discussion designed to acquaint them with the concept of covariation. The training focused on strength and direction as the two key features of covariations between events. The discussion began by the experimenter's pointing out familiar cases in which events tend to covary, asking the subject for examples of his/her own. Possible relationships between events were discussed as being perfectly correlated, imperfectly correlated, or not related at all. Finally, contingencies were discussed in terms of possible direction of relationships, again using familiar examples. No specific strategies were suggested for judging event relationships. The concept training took about 5 min. The no concept training group heard no such introductory discussion.

Table 2
Example Introduction

Some plants need lots of water to stay healthy. For example, they would need a large glass of water each week to stay healthy. However, some plants can be watered too much and will die if they are given a lot of water. These plants would stay healthiest when given a small glass of water each week. Other plants would grow equally well with a big or small glass of water each week. Use these cards to decide whether this plant is one that stays healthiest when given a large glass of water each week or when given a small glass of water each week, or doesn't it make any difference for this plant?

For this plant, a large glass of water is associated with the plant

being healthy	no relationship	dying

Sort instructions vs. no sort instructions. Also of interest to the investigation were the data organization strategies subjects used in judging the covariation problems. Subjects in the sort instructions condition were instructed to sort the data cards into a 2 by 2 matrix before judging each relationship. A matrix was provided, with appropriate row and column labels for each problem. A comparison group of subjects (no sort instructions) were given no instructions about data sorting.

Subjects. Subjects were 80 10th-grade students recruited from a parochial high school and through an ad in the city newspaper. Subjects were randomly assigned to one of the four judgment conditions, with 20 subjects in each condition.

Results

For each subject, response accuracy was summed across problems within strategy type, resulting in a score ranging from 0 to 3 for each of the three problem types. Since the study included repeated measures, a profile analysis was done on the data (McCall & Appelbaum, 1973; Morrison, 1976). Factors included concept training (two levels), sort instructions (two levels), problem sequence (three levels), and repeated measures on problem strategy type (three levels). All interactions between factors were nonsignificant, so we focus our discussion on the main effects.

Judgment conditions. The results indicate that performance was significantly facilitated in the concept training condition. Mean number of correct judgments per problem type was 1.9 for subjects with concept training and 1.6 for people without such training [$F(1,68) = 5.11$, $p < .05$]. Sort instructions, however, did not significantly facilitate judgment accuracy. Mean accuracy was 1.76 for subjects who were instructed to sort and 1.75 for subjects who were not so instructed [$F(1,68) = .037$, $p > .05$]. Any effect here may have been attenuated by the fact that 50% of the subjects sorted the cards spontaneously in the condition in which no sort instructions were given. Problem sequence did significantly influence response accuracy [$F(2,68) = 3.17$, $p < .05$]; mean accuracy was 1.63, 1.65, and 1.99 for sequences beginning with Cell a, sum-of-diagonals, and conditional probability problems, respectively.

Judgment strategy. Problem strategy type significantly influenced performance [$F(2,67) = 42.62$, $p < .01$]. Mean accuracy was 2.27 for Cell a problems, 2.06 for sum-of-diagonal problems, and 1.0 for conditional probability problems. Comparisons between levels suggest that conditional probability problems were significantly more difficult than sum-of-diagonal problems [$F(1,68) = 65.22$, $p < .01$], but sum-of-diagonal problems were not significantly more difficult than Cell a problems [$F(1,68) = 2.29$, $p > .05$].

Most informative about subjects' solution strategies is the analysis of individual subject performance patterns. Subjects were judged to have passed a given level or problem type if they correctly identified the direction of relationship on two out of the three problems of that strategy type. Subjects who judged according to the conditional probability strategy should pass the criterion on all problem types, sum-of-diagonal strategy subjects should pass sum-of-diagonal and Cell a problems, and subjects using the Cell a strategy should pass Cell a problems alone. The probability of meeting each of these criteria by chance is .14 for Cell a, .05 for sum of diagonals, and .02 for the conditional probability strategy. Subjects who passed no criteria were categorized as Strategy 0. All other judgment patterns were labeled unclassifiable.

Categorization of subjects according to response patterns identified five subjects as Strategy 0 (6%), 14 as using Cell a (17%), 33 as comparing sums of diagonal cells (41%), and 14 as comparing conditional probabilities (17%). The remaining 14 subjects (17%) did not conform to any of the patterns. These unclassifiable records were examined for evidence of alternate strategies. Failure to find judgment patterns in common between subjects indicates that any underlying strategies were idiosyncratic. Subjects' response patterns were expected to conform to a Guttman scale. Analysis of the set of individual judgment patterns yielded a coefficient of reproducibility of .93, meeting Guttman's (1967) criterion of scalability.

Discussion

Results indicate that judgment conditions do reliably influence performance on covariation judgments. In particular, performance was improved when subjects participated in a brief discussion about the concept of covariation between events. This significant effect contrasts with Jenkins and Ward's (1965) previous failure to find improved judgments by subjects who solved and received feedback on two pretraining problems. Rather than using problems to train subjects, we involved them in a discussion pointing out key variables (strength and direction) that might be relevant to evaluation of event contingency. Our discussion might have offered information that subjects could not gain from simply solving problems. Admittedly, however, this training effect may have more than one source. The discussion may truly correct misconceptions (or fill knowledge gaps) about relationships between events. Alternatively, subjects exposed to the discussion may simply be more motivated than subjects who heard no such discussion. In fact, our new evidence of subjects' specific judgment rules makes our concept training look very general indeed. Future work on judgment intervention might more profitably be targeted at specific features of underlying judgment strategies.

Our results do, however, suggest that sort instructions are not sufficient to improve covariation judgment. Most impressive is the large proportion (50%) of subjects who sorted the data cards without being instructed to do so. Again, this outcome contrasts with past work (Smedslund, 1963) that found little spontaneous sorting of data instances. Lack of data organization strategies looks like a poor account of past performance failures. Evidence indicates that adolescents will organize the data when the opportunity is available.

The influence of problem sequence on performance was unexpected and, thus, difficult to account for. The best performance was in the order that began with a noncontingent conditional probability problem. The other two orders each began with contingent relationships: one a Cell a problem, the other a sum-of-diagonals problem. Subjects who started with the most difficult problem were most successful, even though they received no feedback on their solutions. This order effect indicates that subjects may be cued into better strategy use simply by the problem-frequency structure of conditional probability problems. However, initial problems in the three sequences differ in both strategy level and nature of relationship. This order effect bears replicating in a design in which these two possible sources are unconfounded.

Of focal interest to the investigation was the possibility of identifying subjects' solution strategies by their judgment patterns across problem types. Results support our proposed judgment rules. First, subject performance indicates that problem difficulty did increase with strategy level. Interpretation of this trend is complicated by the fact that conditional probability problems included only noncontingent relationships. Previous work (Alloy & Abramson, 1978) has indicated that noncontingent problems might be particularly difficult for people. We can, however, restrict our comparison to noncontingent problems in each of the three problem strategy types. This comparison indicates that the problem types are differentially difficult, even within the noncontingent problems [$\chi^2(2) = 23.08$, $p < .01$]. Order of difficulty is the same as that in the overall comparison.

Stronger support for our proposed strategies is the extent to which individual judgment patterns conformed to those predicted. Only 14 of the 80 records failed to match any of the judgment rules. The most common response pattern was congruent with the strategy described by Piaget as formal operational, that is, comparison of the sums of diagonal cells, a + d and b + c. Cell a and conditional probability strategies were about half as likely as this modal strategy.

This rule-diagnostic approach suffers from all the inference problems of strategy modeling in general. That is, a solution pattern conforming to one of the predicted patterns could be the product of an alternative rule that produces judgments isomorphic with the proposed rule. However, congruence with a given strategy does clearly identify the other proposed models as inadequate in characterizing the strategy. Thus, the approach may say more about what a given subject is not using than what he or she is using to identity event contingencies. At the same time, performance patterns on the given problem set severely restrict the realm of possible alternative judgment bases.

Further support for the specific proposed solution rules comes from differentiating the proposed strategies from viable alternatives. In fact, our sum-of-diagonal strategy has a viable competitor that would produce the same judgment pattern. Specifically, a subject who is asked if Outcome A_1 is associated with B_1 or B_2 may simply compare the frequencies in the cells with those particular event combinations ($A_1B_1 - A_1B_2$). With problem structures used in this experiment, the comparison would be between Cells a and b in a traditionally labeled contingency table (a vs. b strategy). Given the specific frequencies chosen for Experiment 1, this a vs. b strategy would result in a solution pattern congruent with our proposed sum-of-diagonals strategy.

Experiment 2 was planned to differentiate among these possible strategies. Table 3 illustrates the actual problem frequencies used to differentiate the four possible strategies. Again, problems are structured hierarchically, so that the conditional probability strategy results in accurate judgment of all problems; the sum-of-diagonals strategy results in accuracy on sum-of-diagonal, a vs. b, and Cell a problems; Strategy a vs. b yields accurate judgment on a vs. b and Cell a problems; the Cell a strategy gives accurate solution to Cell a problems alone.

Experiment 2 offers an opportunity to clarify ambiguities from Experiment 1. Specifically, we were able to identify some contingent problems for the conditional probability strategy, thus making it more comparable to the other problem sets.[2] Also, possible sources of our order effect were unconfounded by controlling direction of relationship and varying problem strategy level of starting problem. This resulted in four problem sequences, each of which began with a noncontingent problem of one of the four strategy types.

Finally, we will be testing an alternative response format, with the hope of eliciting reliable judgments of strength as well as direction of relationship. Although the judged strength of relationship is not critical in this investigation, it will be important to subsequent research into more detailed characterization of judgment strategies.

EXPERIMENT 2

Method

Problem contents included the nine event pairs from Experiment 1, plus three additional problems. The new problems pictured a possible association between space creatures' moods (happy/sad) and the presence or absence of one of three weather conditions (snow, fog, or rain).

Since subjects had such little difficulty with the previous procedure, presentation format for this experiment was modified slightly to permit group testing. As previously, problems were introduced with a paragraph describing a context in which several observations were made on two potentially related variables. For each problem, data instances were pictured in a 2 by 2 table. Subjects were asked to look at the pictured information and identify the relationship between the events. Subjects made their judgments on one of two response forms.

Form A. This response form was an alternative statement of event contingency. The form asked subjects about the relative likelihood of an event (A_1), given each of the two alternative states of the second event (B_1 and B_2). An example question from a problem about space creatures (Blockheads) moods and presence or absence of snow is shown in Table 4. A similar response scale was developed for each problem content. In each case, the question format asked whether A_1 was more, less, or

Table 3

(a) Cell Frequencies Used for Each Problem Type: Experiment 2

Cell a Problems			a vs. b Problems			Sum-of-Diagonal Problems			Conditional Probability Problems		
	B_1	B_2		B_1	B_2		B_1	B_2		B_1	B_2
A_1	11	4	A_1	4	1	A_1	4	4	A_1	2	12
A_2	1	8	A_2	3	16	A_2	1	15	A_2	0	10
	B_1	B_2		B_1	B_2		B_1	B_2		B_1	B_2
A_1	6	6	A_1	4	4	A_1	9	5	A_1	1	5
A_2	6	6	A_2	8	8	A_2	7	3	A_2	3	15
	B_1	B_2		B_1	B_2		B_1	B_2		B_1	B_2
A_1	1	8	A_1	4	11	A_1	4	4	A_1	12	2
A_2	11	4	A_2	8	1	A_2	15	1	A_2	10	0

(b) Strategy Use and Resultant Patterns of Problem Accuracy

Subject Strategy Type	Problem Strategy Type: Cell a	a vs. b	Sum of Diagonals	Conditional Probability
Conditional Probabilities	+	+	+	+
Sum of Diagonals	+	+	+	0
a vs. b	+	+	0	0
Cell a	+	0	0	0
Strategy 0	0	0	0	0

Note—+ = accurate; 0 = inaccurate.

Table 4
Example Question

The picture indicates that when it was snowing blockheads were:

+3	+2	+1	0	−1	−2	−3
much more likely	somewhat more likely	a bit more likely	just as likely	a bit less likely	somewhat less likely	much less likely

to be happy than (as) when it was not snowing. On your answer sheet, write the scale number that best completes the sentence.

Note—The question is taken from a problem about space creatures' (blockheads) moods and presence or absence of snow.

equally likely to occur with B_1 than with B_2. Comparison with problem frequencies in Table 3a should indicate predicated judgments by each of the proposed rules.

Form B. This response form was the same one that was used in Experiment 1.

The 12-problem sequence formed a problem booklet. Problems in the booklets were ordered in one of four sequences, each of which began with a problem of a different strategy level. All initial problems were noncontingent relationships.

Subjects were tested in groups of 18 to 20 people. A brief discussion clarified the nature of the stimuli and judgments to be made. Since data instances were organized into 2 by 2 matrices for each problem, judgment conditions were analogous to the sort instructions/no concept training condition from Experiment 1.

Subjects. Subjects were participants in an introductory psychology class who served in the experiment as one option in fulfilling a course requirement. A total of 184 people participated, 90 responding to Response Form A, 94 to Response Form B.

Results

As in Experiment 1, summary scores for each subject indicated the number of correct judgments on the three problems for each of the four strategy types. A profile analysis was done on the data, including factors of problem sequence (four levels), response form (two levels), and repeated measures on problem strategy type (four levels).

Judgment conditions. Results indicate that the new response form did significantly facilitate judgment accuracy [mean accuracy = 2.29 for Form A, 2.05 for Form B; $F(1,176) = 12.11$, $p < .001$]. As in Experiment 1, problem sequence also influenced accuracy [$F(3,176) = 3.42$, $p < .05$], although the pattern was different from the previous outcome. Mean accuracy by strategy type of initial problem was 2.18 for the Cell a sequence, 2.35 for the a vs. b sequence, 2.03 for the sum-of-diagonal sequence, and 2.12 for the conditional probability sequence. Problem sequence also interacted significantly with problem strategy type [$F(9,528) = 1.96, p < .05$].

Judgment strategy. Judgment accuracy significantly differed as a function of problem type [$F(3,174) = 138.66, p < .001$]. Subjects accurately judged an average of 2.77 of the Cell a problems, 2.57 of the a vs. b problems, 2.04 of the sum-of-diagonal problems, and 1.30 of the conditional probability problems. Performance differences were significant between all pairwise comparisons of adjacent problem types in the strategy hierarchy [$F(1,176) > 16.0$, $p < .001$, for all comparisons).

Individual judgment accuracy patterns were analyzed according to the proposed strategy diagnostic. Again, subjects passed the criterion for each problem strategy type if they were correct on at least two out of the three problems. Table 3b illustrates the predicted strategy patterns. Subjects using the conditional probability strategy should pass all problem types. The sum-of-diagonal strategy should result in correct solution of all problems except the conditional probability type. Subjects using Strategy a vs. b should pass criterion for only a vs. b and Cell a problems. The Cell a strategy should result in correct solution of Cell a problems alone. Again, subjects who passed no criteria were classified as Strategy 0. The probability of meeting these criteria by chance was .08 for Cell a, .03 for a vs. b, .02 for sum of diagonal, and .007 for the conditional probability strategy. Using this scheme, 1% of subjects were classified as using Cell a strategy, 18% as using a vs. b, 35% as using the sum-of-diagonal rule, 33% as comparing conditional probabilities, and 13% as unclassifiable. No subjects were classified as Strategy 0. As in Experiment 1, the unclassifiable records suggested no consistently used alternative strategy. Hierarchical patterns in the set of responses were tested by a Guttman scale analysis. The resultant coefficient of reproducibility was .97, demonstrating close conformity to the Guttman pattern.

Discussion

Experimental results again suggest that judgment conditions influence response accuracy. Particularly interesting is the effect of response form on contingency judgments. Subjects asked to compare the differential likelihood of an outcome under two conditions were more accurate than subjects asked to identify the relationship between the events. Although the two questions were logically equivalent, subjects seem to have interpreted them differently. Further investigation of this difference could help identify specific features of everyday definition of event contingencies. For our own purposes, however, it suffices to note our success in identifying a response form for eliciting reliable judgments about strength as well as direction of relationship. Such information will be important for further investigations of judgment strategies.

Again, problem sequence significantly influenced judgment accuracy, but this time in a pattern different from our previous findings. The conditional probability sequence was no longer the most difficult of the four sequences. We did control for the contingency of initial relationship this time, to clarify our interpretation of the sequence effect. However, variability in the outcome pattern may indicate that there was no reliable effect to interpret.

Most interesting to the investigation is a further definition of underlying judgment strategies. Again, our results indicate differential response accuracy in the hierarchical pattern predicted. Furthermore, individual judgment patterns are also congruent with those predicted by the proposed strategies. Few subjects were characterized by the Cell a strategy (1%). Response patterns match the newly proposed a vs. b strategy for 18% of the subjects. Conditional probability response patterns were common as well (33%). Again, however, the modal strategy was the sum-of-diagonals rule (35%). Thus, Experiment 2 offers a replication of the close congruence between the proposed strategies and subjects' actual judgment patterns.

In overview, the two experiments suggest a set of strategies that characterize adolescent and adult judgment of contingencies between events. Subject's judgment patterns suggest strong intraindividual consistency in rule use. However, the variety of rules evident in our results indicates that characterization of group judgment by any single rule would be inappropriate.

Our rule-analytic approach offers an interpretation of past variability of performance in research in this area. Since faulty judgment rules may produce accurate judgment of a variety of event contingencies, care must be taken in inferring judgment competence from performance on covariation problems. Without problems designed to discriminate between alternative rules, judgment accuracy is an ambiguous index of underlying rule use. The extent of this problem is difficult to evaluate in past studies that fail to report the specific frequencies used (e.g., Inhelder & Piaget, 1958; Neimark, 1975). However, studies that do report cell frequencies (Seggie & Endersby, 1972; Smedslund, 1963) indicate dominant use of problems that would be accurately judged by a rule as simple as our a vs. b strategy. Subjects' performance on problems such as these would be a poor indicator of underlying rule use.

Subjects' judgments of our diagnostic problem set indicates moderate everyday sophistication about event contingencies. Although several 10th-grade subjects judged the relationship according to Cell a frequency alone, college subjects rarely did so. This low frequency contrasts sharply with Smedslund's (1963) suggestion that the strategy is typical among adult judges. However, comparison of frequencies in Cells a and b was a common strategy among college subjects. Essentially, this rule ignores all nonoccurrences of one of the target events in defining the relationship, an approach with severe limitations. The most commonly used strategy is congruent with that proposed by Inhelder and Piaget (1958) as formal operational: comparison of the sums of the diagonal cells in a contingency table. Although a generally successful approach, the rule may result in error when rates of event occurrence and nonoccurrence are disparate for one or both of the variables. The most optimal of the proposed strategies is the comparison of the conditional probabilities $P(A_1/B_1)$ and $P(A_1/B_2)$. This approach was less common among these subjects. Thus, the majority of subjects' judgment patterns indicate judgment rules that are frequently useful, but less than optimal.

Although the subjects in these experiments evidenced some understanding of covariation, level of judgment competence may be specific to judgment conditions. Compared with everyday contexts, subjects in these experiments experienced privileged judgment conditions. Data were organized for the subject (Experiment 2) or were available for the subject to organize (Experiment 1). All relevant information was accessible at the time the judgments were made. In contrast, information about everyday event covariations must be collected and remembered over an extended time span. Relevant information must be selected from a complex decision environment. Thus, subjects' rule use in these experiments may represent an optimistic assessment of everyday competence. Future work could investigate rule use under the memory and attentional demands commonly found in actual decision environments. The more limited rules may be the dominant strategy under such conditions.

The faulty rules evidenced in this experiment have implications for the related process of identifying cause-effect relationships. In particular, several psychologists (Heider, 1958; Inhelder & Piaget, 1958; Kelley, 1967) have suggested that everyday causal judgment is based on a covariation analysis. People may search for likely explanations of everyday events by identifying event covariates as potential causes, but our evidence indicates that faulty judgment would produce systematic errors in that process. As a result, people may see relationships between events that are, in fact, independent, or they may see no relationship between true event covariates. In this way, limited rules for covariation judgment may undermine a person's adequacy at identifying cause-effect relationships.

REFERENCES

Alloy, L., & Abramson, L. Judgment of contingency in depressed and nondepressed students: Sadder but wiser. *Journal of Experimental Psychology: General*, 1979, **108**, 441-485.

Guttman, L. A basis for scaling qualitative data. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence*. New York: Basic Books, 1958.

JENKINS, H., & WARD, W. Judgment of contingency between responses and outcomes. *Psychological Monographs*, 1965, **79**, 1-17.

KELLEY, H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

McCALL, R., & APPELBAUM, M. Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development*, 1973, **44**, 401-415.

MORRISON, D. *Multivariate statistical methods.* New York: McGraw-Hill, 1976.

NEIMARK, E. Longitudinal development of formal operations thought. *Genetic Psychology Monographs*, 1975, **91**, 171-225.

SEGGIE, J., & ENDERSBY, H. The empirical implications of Piaget's concept of correlation. *Australian Journal of Psychology*, 1972, **24**, 3-8.

SMEDSLUND, J. The concept of correlation in adults. *Scandinavian Journal of Psychology*, 1963, **4**, 165-173.

WARD, W., & JENKINS, H. The display of information and the judgment of contingency. *Canadian Journal of Psychology*, 1965, **19**, 231-241.

NOTES

1. For each problem, subjects were subsequently asked to judge the strength of relationship between events. Subjects' frequent confusion about use of this scale led us to drop the strength judgments from further consideration.

2. We had some difficulty defining a noncontingent relationship for the sum-of-diagonals problems. The problem we included (middle problem, Column 3, Table 3) deviates slightly from independence $[P(A_1/B_1) - P(A_1/B_2) = -.06]$ by the conditional probability rule. As a result, we scored responses as correct if subjects concluded that A_1/B_1 was either less likely or just as likely as A_1/B_2. The problem does discriminate appropriately among the other judgment rules. Cell a and a vs. b judges should say that A_1/B_1 is more likely than A_1/B_2; sum-of-diagonal judges should say the two outcomes are equally likely.

(Received for publication December 20, 1979; revision accepted April 9, 1980.)