

Judging interevent contingencies: Being right for the wrong reasons

HARRIET SHAKLEE

Eugene Research Institute, Eugene, Oregon

and

EDWARD A. WASSERMAN

University of Iowa, Iowa City, Iowa

Mathematically inaccurate judgment rules often produce correct covariation judgments; thus, accuracy of covariation judgment alone may be a poor index of the sophistication of a subject's understanding. We offer a past paradigm of our own (Wasserman & Shaklee, 1984) as an instance in which impressive judgment accuracy may have been the product of simple and inaccurate judgment rules. The present investigation replicates the judgment paradigm of our prior experiment, using a set of 12 covariation problems designed to produce unique judgment patterns by each of four judgment rules. Subjects' judgment patterns indicated that use of a mathematically accurate rule was quite rare (comparison of conditional probabilities: 3.1% of subjects). The modal judgment pattern conformed to that predicted by a rule in which subjects compare only two cells of a 2×2 contingency table (Strategy a-versus-b: 38.1% of subjects). Distributions of strategy classifications differed among several judgment conditions which varied in the presentation format of event-frequency information.

Research in human covariation judgment shows poor consensus about people's competence in judging relationships between events. Existing reviews point to several likely sources of these varying levels of covariation judgment competence, including differences in experimental techniques and criteria of accuracy (e.g., Beyth-Marom, 1982; Shaklee, 1983). However, this poor consensus also may stem from a more basic problem of experimental logic which has received little attention in past analyses: namely, accuracy alone may be a poor index of underlying competence at covariation judgment. In particular, correct judgments may be the product of several cognitive rules, some of which represent mathematically incorrect understandings of covariation. Many such rules have been discussed as likely bases of covariation judgment.

A frequently proposed rule (e.g., Inhelder & Piaget, 1958; Shaklee & Tucker, 1980; Smedslund, 1963) is one in which a subject judges a relationship according to the frequency of pairings of the target event states (A_1B_1 , cell a of a contingency table; cell-a rule). For example, the rule would determine a relationship between plant health and plant food to be positive if it is the most frequent combination of the four event-state combinations (i.e., healthy plant-plant food, healthy plant-no plant food, unhealthy plant-plant food, unhealthy plant-no plant food), negative if it is the least frequent combination, and indepen-

dent if it is equal to the other three combinations. By a second simple strategy, a subject may determine whether plant food is associated with plant health by comparing the frequencies of healthy plants treated with plant food to healthy plants not treated with plant food (cells a and b of a traditionally labeled contingency table; strategy a-versus-b). A third approach might judge interevent relations by comparing the event-state combinations that confirm a positive relationship between the target events (healthy plants with plant food, unhealthy plants without plant food; contingency table cells a and d), with those disconfirming the relationship (healthy plants without plant food, unhealthy plants with plant food; contingency table cells b and c). Finally, an individual may use a mathematically accurate rule to judge interevent covariation. One frequently proposed rule compares the probabilities of the target event (e.g., plant health) given each of the alternative states of the other event (plant food/no plant food; conditional probability strategy).

All of these proposed rules incorporate at least some of the information relevant to event covariation. As a result, even the simplest rule will lead to accurate judgment of some covariation problems. However, other problems can be constructed which would be incorrectly judged by all rules except the conditional probability rule (see Shaklee, 1983, for a more detailed discussion of these rules and their likely accuracy). By this analysis, the relationship between judgment accuracy and strategy is indeterminate. Accurate judgment is not in itself sufficient ground to infer mathematical sophistication at covariation judgment.

Reprint requests should be sent to H. Shaklee, Eugene Research Institute, Eugene, OR 97401.

This indeterminacy might not be such a concern if researchers tended to employ problems that could be accurately judged by only the most sophisticated of the rules. However, a review of past problem sets indicates the contrary. Seggie and Endersby (1972) and Smedslund (1963) each used problems for which accurate directional judgments could be made by a rule as simple as the a-versus-b strategy. Ward and Jenkins (1965) used problems for which subjects' judgments about relationship strength were intercorrelated with all of their proposed judgment rules (most r s were greater than .50 and some were as high as .80). The extent of the problem cannot be assessed in studies which do not report the event-state frequencies used (e.g., Inhelder & Piaget, 1958). This is especially true of paradigms in which subjects are free to respond in order to determine their control over an outcome; in these studies individual subjects generate unique frequencies of event-state combinations, depending on their patterns of response choices (e.g., Allan & Jenkins, 1980; Alloy & Abramson, 1979). Evidence of judgment accuracy in such paradigms is uncertain grounds for inferring judgment competence.

A few experiments have included problems that require a conditional probability rule for accurate solution. However, in each case, those problems were carefully constructed to be erroneously judged by other proposed rules (Shaklee & Tucker, 1980 and subsequent studies; Allan & Jenkins, 1983; Arkes & Harkness, 1983). Considered in the context of past research, our conclusion is that the experimenter must make a special effort to generate problems that require accurate covariation rule use for correct judgment. Without such an effort, a nondiagnostic problem set may result.

We offer a problem set of our own as a case study (Wasserman & Shaklee, 1984). Subjects were asked to judge 24 different problems about the effect of a response (tapping a wire of a radio) on an outcome (a buzz in the radio). Results of several experiments showed that subjects' ratings of contingency were quite sensitive to the true contingency represented in the problems. Congruence was particularly close in conditions where the tap-buzz information was presented in a tabled format. Here, group mean judgment of relationship strength increased monotonically as a function of the true strength of the relationship, crossing the zero point on the rating scale at the point of true noncontingency. Furthermore, strength ratings were symmetrical about zero, with positive relationships receiving the same strength ratings as negative relationships of the same true strength.

In all of these respects, the results suggest that subjects were closely attuned to true contingency in rating response-outcome relations. However, further analysis of the problem set discloses a notable difficulty: accurate judgment of each problem can be the product of a mathematically inaccurate rule. The correlation between conditional probability and cell-a judgments on the problem set is modest ($r = .42$), but the correlation with the other

two rules is substantial ($r = .78$ with a-versus-b; $r = .87$ with the sum of diagonals rule).

One way to address the question of strategy use is to construct a problem set designed to produce a distinctive judgment pattern by each of the proposed judgment rules. Shaklee and colleagues have successfully used this rule-analytic approach to identify covariation judgment strategies in a variety of studies (e.g., Shaklee & Hall, 1983; Shaklee & Tucker, 1980). Results of these studies indicate that a minority of college subjects were categorized as using the conditional probability rule. In each case, sizable proportions of subjects produced judgment patterns congruent with a-versus-b and sum of diagonals rules. Such evidence indicates that the high level of accuracy seen in our recent troubleshooting paradigm may indeed be the product of suboptimal judgment strategies.

We designed the present study to analyze the strategies that subjects employ in this paradigm. The earlier problem set does not permit a strategy analysis, since correct judgments can be made with the a-versus-b, sum of diagonals, and conditional probability rules. However, we can employ a set of problems in the prior paradigm that yield distinctive judgment patterns by each of the four proposed rules. In the present experiment, subjects judged such a rule-diagnostic problem set under the same information conditions as those in Wasserman and Shaklee (1984, Experiment 4).

METHOD

Subjects

The subjects were 160 students in introductory psychology classes (20 males and 20 females in each of four judgment conditions).

Problems

Subjects judged 12 different problems about the relationship between the tapping (T) of a wire and the buzzing (B) of a radio. This set, shown in Table 1, included three problems of each of four types. Cell-a problems would be accurately judged by all four strategies. The a-versus-b problems would be accurately solved by all strategies except the cell-a rule. The sum of diagonals problems would be accurately judged by sum of diagonals and conditional probability rules.¹ And the conditional

Table 1
Strategy Diagnostic Problem Set

Strategy Type	Relationship Direction	Frequencies of Response-Outcome Pairings in Each Problem			
		Tap-Buzz	Tap-No Buzz	No Tap-Buzz	No Tap-No Buzz
Cell-a	+	11	4	2	7
	0	6	6	6	6
	-	2	7	11	4
a-versus-b	+	7	3	2	12
	0	3	3	9	9
	-	4	11	8	1
Sum of Diagonals	+	2	2	2	18
	0	9	5	7	3
	-	8	8	8	0
Conditional Probability	+	10	12	0	2
	0	1	3	5	15
	-	12	10	2	0

probability problems could be accurately solved by the conditional probability rule alone. The problem set included one each of noncontingent, positive, and negative relationships for each of the four strategy types. The contingent problems each represented relationships of moderate strength [$p(B/T) - p(B/no T) = \pm .40$ to $.62$].

Procedure

Subjects were asked to take the role of a person who, in troubleshooting a malfunctioning radio, suspects that a loose wire may be the cause of an occasional buzzing sound. The individual taps the wire to see if it has any effect on the radio's buzz. Subjects judged the tap-buzz relations under one of four conditions (detailed descriptions of the conditions are available in Wasserman & Shaklee, 1984, Experiment 4):

Line. Response-outcome intervals were represented by dashes on a time line; taps were represented by an A above the time line, and buzzes were represented by a B below the time line.

Line-interval. Tap-buzz information was represented as in the line condition, but subjects were trained to segment the time line into response-outcome intervals.

Line-table. Problems were represented in time lines as above, and subjects were taught to segment the lines into response-outcome intervals. In addition, subjects were asked to count the frequencies of each of the four types of response-outcome pairs (i.e., tap-buzz, tap-no buzz, etc.) and to enter that information into the table provided.

Table. Response-outcome frequency information was presented in tables like those above, with the completed blanks showing the correct frequencies of the response-outcome pairs.

For each problem, subjects judged the effect of tapping on the wire (from -4 , prevents sound from occurring, to $+4$, causes sound to occur).

RESULTS

Each subject was categorized by judgment strategy according to his or her accuracy profile on problems of the four strategy types. A subject was deemed to have passed criterion on problems of a particular type if he or she correctly judged two or more of the three problems of the type. Subjects who passed criterion on all of the problem types were categorized as using the conditional probability rule. Subjects who passed all problem types except conditional probability were classified as sum of diagonals judges. Those who accurately judged only the cell-a and a-versus-b problems were labeled a-versus-b judges. Cell-a judges accurately solved cell-a problems alone. Subjects who passed criteria on none of the problem types were called Strategy 0. Those who passed problem type criteria in some other pattern were unclassifiable.

Table 2 shows the strategy classifications of subjects in each of the four judgment conditions. Comparison of the distributions shows them to be reliably different [$\chi^2(15) = 53.60, p < .01$]. The two table conditions

produced relatively high rates of a-versus-b judgment patterns (65.0% in line-table, 50.0% in table). Sum of diagonals and conditional probability judgment patterns were rare in all conditions (sum of diagonals: 5.0% to 12.5%; conditional probability: 0.0% to 7.5%). Unclassifiable and Strategy 0 subjects were more common in the two line conditions (37.5% and 45.0% for Unclassifiable and 17.5% and 20.0% for Strategy 0 in line-interval and line conditions, respectively). Cell-a judgment patterns were most common in the table condition (22.5%).

DISCUSSION

The conceptual analysis of our original troubleshooting problem set indicated that subjects could judge many of those problems accurately by a strategy as simple as the a-versus-b rule. The present empirical analysis indicates that the a-versus-b rule actually was the most likely source of judgment patterns in this paradigm. Across the four judgment conditions, a-versus-b was by far the most common judgment pattern shown (38.1% of the subjects), and it was also the modal category of classifiable strategies in each of the four judgment conditions. We also note that strategy use differed among judgment conditions, with a-versus-b rule classifications especially prominent in the table and line-table conditions (50.0% and 65.0%, respectively). Strategy classifications in these conditions contrasted sharply with those in the line and line-interval conditions, where substantial numbers of judgment patterns failed to conform to any of our rules (65.0% Unclassifiable plus Strategy 0 in line; 55.0% in line-interval). These indeterminate judgment patterns may have several possible sources, including random responding, shifting rules at some point in the problem set, or systematic use of another rule. However, the most likely interpretation stems from the presentation format of the frequency information. Subjects in the table condition were given numerical summaries of the event-state combinations; line-table subjects generated these summaries on their own. Such tallies would be easy to combine in some systematic fashion to judge the problems. Subjects who had no comparable numerical summaries might tend to estimate rather than to count the relevant frequencies. Errors in these estimates could result in judgment errors, even by the best of judgment rules. Such errors would tend to be unpredictably sprinkled throughout the problem set. In this way, someone systematically using one of our proposed rules might generate an accuracy pattern on the problem set which would be unclassifiable by our rule-analytic scheme.

Evidence of subjects' strategy use in this troubleshooting paradigm underscores the problem of inferring judgment sophistication solely from the overall accuracy of contingency ratings. The close congruence in Wasserman and Shaklee (1984) between subjects' judgment functions on the 24-problem set with the true problem contingencies was impressive in view of the range of contingencies represented. However, our strategy analysis revealed that use of a mathematically sophisticated rule of contingency was rare in this judgment paradigm. Only 3.1% of the subjects were categorized as using the optimal conditional probability rule and 8.1% as using the sum of diagonals strategy. Evidence of cell-a and a-versus-b rule use was much more likely.

Table 2
Strategy Classification Frequencies in Each Judgment Condition
(Percentages in Parentheses)

Judgment Condition	Unclassifiable	Strategy 0	Cell-a	a-versus-b	Sum of Diagonals	Conditional Probability
Table	4 (10.0)	1 (2.5)	9 (22.5)	20 (50.0)	5 (12.5)	1 (2.5)
Line-Table	2 (5.0)	2 (5.0)	7 (17.5)	26 (65.0)	3 (7.5)	0 (0.0)
Line-Interval	15 (37.5)	7 (17.5)	4 (10.0)	10 (10.0)	3 (7.5)	1 (2.5)
Line	18 (45.0)	8 (20.0)	4 (10.0)	5 (12.5)	2 (5.0)	3 (7.5)
All	39 (24.4)	18 (11.2)	24 (15.0)	61 (38.1)	13 (8.1)	5 (3.1)

Comparing these strategy classifications with Shaklee's prior work shows some striking differences. In particular, the same rule-analytic technique in other work with college students has found substantially more sophisticated rule use than was observed here under analogous conditions. The research of Shaklee and Tucker (1980, Experiment 2) and Shaklee and Hall (1983) entailed comparable paradigms, with event-state frequencies represented in tabled format similar to the table condition in the present study. Categorization of subjects' judgments in the two earlier studies showed cell-a judgment patterns to be rare in each (1.0% and 5.2%). Classifications of a-versus-b rule use were common in each (18.0% and 36.2%), but not as high as the 50.0% found in the present table condition. Sum of diagonals rule use in the present study (12.5%) was close to the range found in past research (35.0% and 15.5%). However, conditional probability rule use was most notably different in the two paradigms, with only 2.5% of the present subjects in the table condition using this rule compared to 33.0% and 31.9% so classified in the previous studies. Comparing the two approaches indicates two likely sources of these differences: response format and stimulus format.

A comparison of response format in the two paradigms indicates that the present study asked subjects to make a causal judgment (response causes, prevents, or has no effect on outcome), whereas the prior studies asked subjects to make a judgment of differential likelihood (outcome more likely given response, no response, or no difference). Two aspects of these formats may affect strategy selection. First, the differential likelihood response format is a verbal expression of a comparison of conditional probabilities (i.e., A more likely given B₁, B₂, or no difference). It would make sense that this response format would be most likely to elicit conditional probability rule use. Beyth-Marom (1982) offers persuasive evidence that subjects are responsive to the particular terms chosen in scales of covariation judgment. Second, Adi, Karplus, Lawson, and Pulos (1978) have found that subjects use less sophisticated strategies when judging causal relationships than when making covariation judgments. Past theories that people select event covariants as causes of events would imply that everyday rules of event covariation (however flawed) should take the same form as the strategies people use to judge cause-effect relationships (Inhelder & Piaget, 1958; Kelley, 1967). The differences in the strategies that subjects use to make covariation and causal judgments in related paradigms may indicate that lay decision makers do not share theorists' belief in the close link between the states of causation and covariation.

Alternatively (or additionally), the critical difference between our present and past findings may lie in the presentation format. The current format of tabled information listed the contingency table cells vertically—from cells *a* to *d*. Since cells *a* and *b* are the first two and they are immediately adjacent to each other, this format may invite comparisons between these two cells. In contrast, the 2×2 table used in Shaklee's prior studies may enhance comparisons among all four cells, since each is adjacent (vertically, horizontally, or diagonally) to the other three. In this way, the presentation format of event-frequency information may differentially affect the ease with which a subject applies alternative judgment strategies.

In overview, our research offers a case study in the relationship between accuracy and strategy in covariation judgment. A subject who judges an interevent contingency incorrectly is certainly doing something wrong. However, an accurate judgment about the relationship may

also be the product of an erroneous judgment rule. For this reason, judgment accuracy alone may be a poor index of a subject's understanding of event covariation.

REFERENCES

- ADI, H., KARPLUS, R., LAWSON, A., & PULOS, S. (1978). Intellectual development beyond elementary school: VI. Correlational reasoning. *School Science & Mathematics*, *78*, 675-683.
- ALLAN, L. G., & JENKINS, H. M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology*, *34*, 1-11.
- ALLAN, L. G., & JENKINS, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning & Motivation*, *14*, 381-405.
- ALLOY, L. B., & ABRAMSON, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*, 441-485.
- ARKES, H. R., & HARKNESS, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, *112*, 117-135.
- BEYTH-MAROM, R. (1982). Perception of correlation revisited. *Memory & Cognition*, *10*, 511-519.
- INHELDER, P., & PIAGET, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- KELLEY, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15). Lincoln, NE: University of Nebraska Press.
- SEGGIE, J., & ENDERSBY, H. (1972). The empirical implications of Piaget's concept of correlation. *Australian Journal of Psychology*, *24*, 3-8.
- SHAKLEE, H. (1983). Human covariation judgment: Accuracy and strategy. *Learning & Motivation*, *14*, 433-448.
- SHAKLEE, H., & HALL, L. (1983). Methods of assessing strategies for judging covariation between events. *Journal of Educational Psychology*, *75*, 583-594.
- SHAKLEE, H., & TUCKER, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition*, *8*, 459-467.
- SMEDSLUND, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, *4*, 165-173.
- WARD, W. C., & JENKINS, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, *19*, 231-241.
- WASSERMAN, E. A., & SHAKLEE, H. (1984). Judging response-outcome relations: The role of response-outcome contingency, outcome probability, and method of information presentation. *Memory & Cognition*, *12*, 270-286.

NOTE

1. We had some difficulty defining a noncontingent relationship for the sum of diagonals problems. The problem we included (see Table 1) deviates just slightly from independence [$p(B/T) - p(B/\text{no } T) = -.06$] by the conditional probability rule.

(Manuscript received for publication October 16, 1985.)