COGNITIVE PSYCHOLOGY 26, 209–239 (1994)

The Accuracy of Intuitive Judgment Strategies: Covariation Assessment and Bayesian Inference

CRAIG R. M. MCKENZIE

University of Chicago

Most of people's apparent strategies for covariation assessment and Bayesian inference can lead to errors. However, it is unclear how often and to what degree the strategies are inaccurate in natural contexts. Through Monte Carlo simulation, the respective normative and intuitive strategies for the two tasks were compared over many different situations. The results indicate that (a) under some general conditions, all the intuitive strategies perform much better than chance and many perform surprisingly well, and (b) some simple environmental variables have large effects on most of the intuitive strategies' accuracy, not just in terms of the number of errors, but also in terms of the kinds of errors (e.g., incorrectly accepting versus incorrectly rejecting a hypothesis). Furthermore, common to many of the intuitive strategies is a disregard for the strength of the alternative hypothesis. Thus, a key to better performance in both tasks lies in considering alternative hypotheses, although this does not necessarily imply using a normative strategy (i.e., calculating the ϕ coefficient or using Bayes' theorem). Some intuitive strategies take into account the alternative hypothesis and are accurate across environments. Because they are presumably simpler than normative strategies and are already part of people's repertoire, using these intuitive strategies may be the most efficient means of ensuring highly accurate judgment in these tasks.

How well do we assess relations between variables, estimate the likelihood of events, test hypotheses, or update our beliefs in light of new evidence? When faced with such judgment tasks, people often appear to use strategies that are simpler and therefore less accurate than formal strategies (see, e.g., Dawes, 1988; Einhorn & Hogarth, 1981; Hogarth, 1987, 1990; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1979; Nisbett & Ross, 1980; Slovic, Fischhoff, & Lichtenstein, 1977). However, although it is clear that subjects' intuitive strategies *can* lead to

Portions of this research were presented at the 1991 Meeting of the Judgment and Decision Making Society, and at the 1992 ORSA/TIMS Joint National Meeting, both in San Francisco, CA. For their helpful comments, I thank Sema Barlas, Brian Gibbs, Gerd Gigerenzer, William M. Goldstein, Alan Hartley, Jonathan J. Koehler, James Shanteau, participants at the University of Chicago's Decision Research Workshop, several anonymous reviewers, and especially Robin Hogarth and Joshua Klayman. Address correspondence and reprint requests to the author at the Center for Decision Research, University of Chicago, 1101 E. 58th Street, Chicago, IL 60637. Email: fac_mckenzie@gsbvax.uchicago.edu.

0010-0285/94 $9.00

errors, it is unclear how often and to what degree these strategies lead to errors in natural contexts.

Through Monte Carlo simulation, the present article investigates the accuracy of several intuitive strategies for covariation assessment and Bayesian inference under some general conditions, and characterizes some conditions under which most of the intuitive strategies perform particularly well or badly. In addition, a common shortcoming of many of the intuitive strategies is pointed out, and a general prescription is provided for how to deal effectively with both tasks.

Examining the accuracy of intuitive judgment strategies is important for at least four reasons. First, it is not clear which strategies are generally accurate and which are not. For certain strategies, it is sometimes argued—and more often just assumed—that they are generally accurate (e.g., availability; Tversky & Kahneman, 1973, 1974). For other strategies, however, it is claimed (or strongly implied) that only chance-level performance (or worse) will be attained through their use. Examples include confirmatory hypothesis testing (Wason, 1960, 1968; Wason & Johnson-Laird, 1972) and making judgments based on "pseudodiagnostic" information (Doherty, Mynatt, Tweney, & Schiavo, 1979). For still other strategies, it is largely unclear how one will fare by using them. Examples here include underweighting or ignoring particular sources of information in covariation assessment (Jenkins & Ward, 1965; Smedslund, 1963; Ward & Jenkins, 1965; Wasserman, Dorner, & Kao, 1990), "averaging" old and new information in belief updating (Anderson, 1981; Hogarth & Einhorn, 1992; Lopes, 1985, 1987; Shanteau, 1970, 1972, 1975), and ignoring base rates when estimating probabilities (e.g., Bar-Hillel, 1980; Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1974).

Second, if it is the case that a simple judgment strategy is generally accurate, this would constitute a functional explanation as to why people use the strategy. Intuitive strategies are assumed to be the result of cognitive economy (Simon, 1955, 1956, 1981; Tversky & Kahneman, 1974) and, therefore, imperfectly accurate. What one would expect to find in a well-adapted system, then, are strategies that are simple and generally accurate. Indeed, there is evidence in the choice-strategy literature indicating that people's simple strategies are often accurate (Payne, Bettman, & Johnson, 1988, 1990; Russo & Dosher, 1983; Thorngate, 1980). As mentioned, though, there is little evidence regarding the accuracy of intuitive judgment strategies, thereby limiting functional explanation in the area of probabilistic judgment.

Third, learning about the accuracy of intuitive strategies can help improve judgment. For example, more emphasis should be placed on avoiding a strategy that often leads to large errors than on avoiding one that is virtually always correct. Furthermore, it would be useful to identify and exploit simple strategies whose performance virtually equals that of a normative counterpart—if such strategies exist. Normative strategies are often difficult to implement, whereas intuitive strategies are presumably simpler and already part of people's repertoire. Thus, prescribing robust intuitive strategies rather than normative strategies may sometimes be a more efficient means of improving judgment.

Last, inextricably tied in with the above three points is understanding the environmental conditions that affect the accuracy of intuitive strategies. Some intuitive strategies may perform well under certain conditions, but poorly under others. If so, this would highlight the importance of examining the relation between environment and judgment strategy. Conclusions about the accuracy of any intuitive judgment strategy may be incomplete—even wrong—without understanding the environmental conditions under which the strategy is used.

The article is divided into three sections. The first two sections compare the intuitive and normative strategies for covariation assessment and Bayesian inference, respectively. For each task, I present simulation results showing the accuracy of several strategies under some general conditions, as well as how some simple environmental factors have a large impact on accuracy. In the last section, I discuss the implications of the results for understanding intuitive judgment strategies and point out opportunities for future research. In addition, I note a commonality between many of the intuitive strategies for both tasks: Subjects often do not take into account the strength of the alternative hypothesis. Finally, I point out that a key to better performance in both tasks lies in considering the alternative, although this does not necessarily imply using a normative strategy (i.e., calculating the ϕ coefficient or using Bayes' theorem). Some intuitive strategies take into account the alternative hypothesis and are accurate across environments.

COVARIATION ASSESSMENT

As pointed out by other researchers, people's intuitive ability to assess the covariation between variables plays an important role in many areas of research in psychology. Examples include learning (e.g., Hilgard & Bower, 1975), attribution (Kelley, 1967), judgments of causality (Einhorn & Hogarth, 1986), decision making (Seggie & Endersby, 1972), clinical assessment (Chapman & Chapman, 1967, 1969; Smedslund, 1963), implicit personality theories (Bruner & Tagiuri, 1954), stereotyping (Hamilton 1976; Hamilton & Gifford, 1976), scientific reasoning (Mynatt, Doherty, & Tweney, 1977, 1978), categorization (Smith & Medin, 1981), and helplessness and control (Seligman, 1975). Indeed, Crocker (1981) emphasizes that accurate covariation assessment enables one to explain the past, control the present, and predict the future.

In most covariation tasks, there are two variables (Event and Outcome) that are present or absent, thereby creating a 2 × 2 contingency matrix (see Figure 1). Subjects are typically asked to assess the direction and/or strength of the relation between the variables, given the four cell values.

Imagine, for example, that you are asked to assess the degree to which a treatment (Event) and recovery from an illness (Outcome) are related, given the following information: (a) 5 instances of treatment followed by recovery, (b) 15 instances of treatment followed by no recovery, (c) 5 instances of no treatment followed by recovery, and (d) 25 instances of no treatment followed by no recovery. How would you use this information to determine the relation between the treatment and recovery? One normative strategy for assessing the covariation between two dichotomous variables is to calculate the ϕ coefficient, $(AD - BC)/[(A + B)(C + D)(A + C)(B + D)]^{1/2}$, where A, B, C, and D represent the respective cell values. The ϕ coefficient is a special case of Pearson's product–moment correlation coefficient, varying from -1 to 1.

Research examining how people assess the covariation between two dichotomous variables has uncovered several strategies that subjects appear to use. Relative to normative standards, virtually all intuitive strategies lead to errors (for reviews, see Alloy & Tabachnik, 1984; Crocker, 1981; Nisbett & Ross, 1980). There is little evidence, however, regarding the degree to which these strategies are inaccurate and under what conditions.

The plan of this section is to describe one normative and seven intuitive covariation strategies, present results showing how well the intuitive

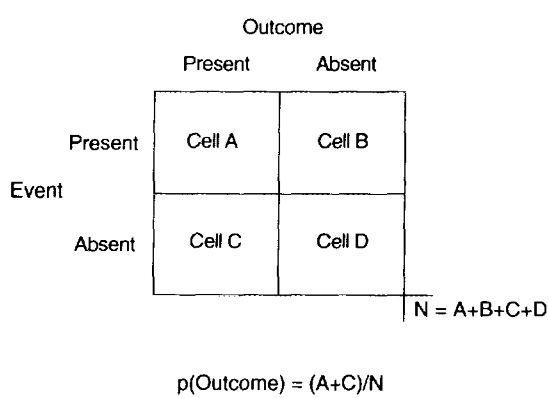

FIG. 1. The four cells of a 2 × 2 contingency matrix.

strategies perform under some general conditions, and characterize some environmental conditions that have a large impact on most of the strategies.

Covariation strategies

ϕ. The ϕ coefficient is the normative covariation measure used in this section (see formula, above). Although there are other normative measures, such as χ^2, that could have been used, ϕ was chosen because it captures both the size and direction of a relation, whereas χ^2 measures only the size.

ΔR. This strategy was proposed as normative by Jenkins and Ward (1965) and is the only intuitive strategy considered correct by covariation researchers. It involves examining the difference between the two row conditional probabilities, that is, $[A/(A + B)] - [C/(C + D)]$ (e.g., the difference between the probability of recovery given treatment and the probability of recovery given no treatment). There is evidence that some subjects use this strategy to judge covariation (Arkes & Harkness, 1983; Shaklee & Mims, 1981, 1982; Shaklee & Tucker, 1980; Ward & Jenkins, 1965; Wasserman et al., 1990).[1] Although the strategy was proposed as normative, Allan (1980) shows that under certain circumstances ϕ will show a change in relation between two matrices while ΔR will not. Because there are questions concerning the normative status of ΔR, it is labeled solely as intuitive for the purposes of this article.

Sum of diagonals. When using this strategy, $(A + D) - (B + C)$, the larger the difference between evidence for a positive relation (Cells *A* and *D*) and evidence for a negative relation (Cells *B* and *C*), the stronger the relation between the variables. Although this strategy was put forth as normative by Inhelder and Piaget (1958), Jenkins and Ward (1965) showed that when either the row or column marginals are not equal, the measure may assess the direction of a relation incorrectly. Nonetheless, there is evidence that some subjects use this strategy (Arkes & Harkness, 1983; Shaklee & Mims, 1981, 1982; Shaklee & Tucker, 1980; Wasserman et al., 1990).

Positive testing. This intuitive strategy was proposed by Klayman and Ha (1987) and involves examining both Row 1 (i.e., the probability of recovery given treatment), and Column 1 (i.e., the probability of treat-

[1] All of the intuitive strategies discussed in this article are algebraic. However, no commitment is made to the cognitive processes involved, only that there is evidence that subjects respond "as if" they are making calculations suggested by the models (see Hoffman, 1960).

ment given recovery), corresponding to $[A/(A + B)] + [A/(A + C)]$. McKenzie (1993) provides evidence showing subjects' preference for both Row 1 and Column 1 in a hypothetical covariation task.

Proportion of hits. Recent evidence indicates that subjects are mostly concerned with Cells A and B when judging covariation (McKenzie, 1993; Wasserman et al., 1990). One use of this information is to examine $A/(A + B)$, the probability, for example, of recovery given treatment. Evidence that some subjects use this strategy is supplied by Ward and Jenkins (1965).

Hits minus false positives. Another use of Row 1 information is to calculate the difference between Cells A and B (i.e., $A - B$). Arkes and Harkness (1983) and Wasserman et al. (1990) have collected data indicating that some subjects use this strategy.

Positive hits. Several researchers have found that some subjects base their covariation judgments almost entirely on Cell A, the joint presence of the variables (Arkes & Harkness, 1983; Jenkins & Ward, 1965; Shaklee & Mims, 1982; Shaklee & Tucker, 1980; Smedslund, 1963; Ward & Jenkins, 1965).

Aggregate model. When averaging across subjects instead of looking at individual subjects' strategies, Cell A is used most heavily in covariation judgments, followed by Cells B, C, and D, in order. These across-subjects results have been found when subjects' responses were regressed onto cell frequencies (Schustack & Sternberg, 1981) and when subjects were asked directly about which cells are important in their judgments (Crocker, 1982; McKenzie, 1993; Wasserman et al., 1990). Furthermore, in a meta-analysis of covariation assessment, Lipe (1990) found that all four cells had an impact on subjects' judgments and were used in the right direction (i.e., Cells A and D were used as evidence for a positive relation and Cells B and C evidence for a negative relation; see also Schustack & Sternberg, 1981; Wasserman et al., 1990). For the purposes of the simulation, the differential impact of the cells is represented in a weighted model as $4A - 3B - 2C + D$.

The above eight strategies (one normative and seven intuitive) form the basis of analysis in this section.

Method

A computer program generated every 2×2 matrix based on all combinations of cell values between 1 and 50. That is, the four cell values were varied independently between 1 and 50, producing 50^4 (6.25 million) matrices. Each covariation strategy's output was calculated for each matrix using the above formulas. Two measures of accuracy were used. The first was each intuitive strategy's product–moment correlation with ϕ. The second was how often each intuitive strategy correctly assessed the direction of the relation. Because there

TABLE 1
Correlations between the Covariation Strategies' Outputs under the General Conditions

Strategy	2	3	4	5	6	7	8
1. ϕ	.99	.95	.82	.70	.67	.47	.86
2. ΔR	—	.94	.81	.71	.67	.47	.86
3. Sum of diagonals		—	.77	.67	.71	.50	.91
4. Positive testing			—	.86	.82	.77	.91
5. Proportion of hits				—	.94	.67	.85
6. Hits minus false positives					—	.71	.90
7. Positive hits						—	.73
8. Aggregate model							—

is no straightforward way to determine direction for the strategy Positive Hits, it is not included in the latter analysis.[2]

Results and Discussion

Performance under the general conditions. Table 1 shows the product-moment correlation matrix for the eight strategies' outputs based on the 6.25 million cases. Most important is the first row, which shows that the correlations between ϕ and the intuitive strategies range from .47 for Positive Hits to .99 for ΔR. The two strategies using only Row 1 information, Proportion of Hits and Hits Minus False Positives, have correlations with ϕ of .70 and .67, respectively. This is of interest because using only Row 1 information appears to be the most common intuitive strategy (McKenzie, 1993; Wasserman et al., 1990). Therefore, these two strategies might be said to represent the "modal subject." Aggregate Model (the "mean subject") correlates .86 with ϕ. Also of interest are rows 2 through 7 of Table 1, which show that the correlations between the seven intuitive strategies range from .47 to .94.

The above results show the correlations between the strategies' strength-of-relation responses. Another relevant issue is how often the intuitive strategies accurately assess the *direction* of a relation. The percentage of times each strategy was correct is shown in the middle column of Table 2. The lowest is 75.0, associated with "modal subject" strategies that use only two cells. Also important is that the "mean subject" Ag-

[2] Instead of creating a third category, the cases in which the strategies imply no relation are defined as implying a positive relation. (The consequences of this simplification are minimal; for example, less than .2% of the matrices result in ϕ equal to zero.) With two exceptions, the strategies were defined as implying a negative relation if the output was less than zero and a positive relation otherwise. The boundary points for Positive Testing and Proportion of Hits were 1 and .5, respectively.

TABLE 2
Percent Correct of Directional Relation for Various Intuitive Covariation Strategies under Different Conditions

Strategy	All matrices	$\|\phi\| \geq .2$
ΔR	100.0	100.0
Sum of diagonals	92.0	99.0
Positive testing	80.4	91.7
Proportion of hits	75.0	84.7
Hits minus false positives	75.0	84.7
Aggregate model	84.2	94.9

Note. The strategy Positive Hits is not shown because there is no straightforward way to determine direction.

gregate Model is accurate 84.2% of the time. Finally, note that ΔR, generally accepted by covariation researchers as normative, is always correct.[3]

Are the directional errors "big" ones? Many of the directional errors may occur either when ϕ is small and positive and the intuitive strategies imply a small negative relation, or vice versa. Such "small" errors can be contrasted with "big" ones, that is, falsely believing that a strong relation does or does not exist. One way to investigate this possibility is to calculate the accuracy of the strategies for only those matrices for which the absolute value of the ϕ coefficient exceeds some value. This would indicate how likely it is that a strategy will miss reasonably strong relations that really exist. To this end, the subset of matrices for which ϕ was either less than or equal to $-.2$ or greater than or equal to .2 was examined. Although the value .2 is arbitrary, there is evidence that subjects are largely unable to detect covariations of this magnitude (Jennings, Amabile, & Ross, 1982), thereby making an error of this size at least psychologically small. For these 3,472,994 matrices, the number of times that each intuitive strategy accurately assessed the direction was calculated. The results are presented in the right column of Table 2 and show that eliminating the matrices implying a weak relation substantially improved the strategies' performance.

Environmental conditions that affect performance. The general analysis involved every possible matrix (given the constraint on cell size). However, this may not correspond to environmental conditions in general, nor to a given environment of interest. For example, in the physi-

[3] These analyses were also performed using all matrices with cell values between 1 and 10. The correlations and percentages correct are virtually identical to those reported here using cell values between 1 and 50. In addition, Spearman rank-order correlations were computed and are virtually identical to the product–moment correlations.

cian's environment of assessing covariations between diseases and symptoms, Cell *D* will usually be larger than the other three cells (i.e., most people have neither the disease nor the symptom). Thus, it is useful to characterize some environmental conditions under which the intuitive strategies perform particularly well or badly.

For four of the strategies, Cells *A* and *B* have more impact on judgments than do Cells *C* and *D*. That is, subjects often prefer information about the Outcome when the Event occurs over when it does not occur. In the case of Proportion of Hits and Hits Minus False Positives, Cells *C* and *D* are completely ignored. For both Positive Testing and Aggregate Model, Cells *A* and *B* also have bigger impact than do Cells *C* and *D;* Positive Testing weights Cell *A* most and ignores Cell *D*, and Aggregate Model weights Cells *C* and *D* least. Because of this relative insensitivity to Cells *C* and *D,* errors are most likely to occur when the Outcome is very common or uncommon. Using the "treatment and recovery" example, if the frequency of recovery is high, the likelihood of mistakenly believing in a positive relation between treatment and recovery increases. If the frequency of recovery is low, however, the likelihood of mistakenly believing in a negative relation increases. (For empirical evidence of this, see Allan and Jenkins, 1980, 1983; Alloy and Abramson, 1979; Jenkins and Ward, 1965; Wasserman and Shaklee, 1984.)

In order to investigate the effect of p(Outcome) on the performance of the strategies, three subsets of the 6.25 million matrices were examined: Those matrices for which p(Outcome) (i.e., $[A + C]/N$) was (a) less than or equal to .1 (34,080 matrices), (b) between .45 and .55, inclusive (1,663,424 matrices), and (c) greater than or equal to .9 (34,080 matrices).

The correlations between ϕ and the intuitive strategies when p(Outcome) is near .50 are shown in the middle column of Table 3. Virtually all the strategies perform well. The correlations when p(Outcome) is either

TABLE 3
Correlations between ϕ and the Intuitive Covariation Strategies' Outputs for Different p(Outcome)

Intuitive strategy	p(Outcome)	
	≥.45 and ≤.55	≤.1 or ≥.9
ΔR	.99	.97
Sum of diagonals	.97	.74
Positive testing	.89	.41
Proportion of hits	.94	.14
Hits minus false positives	.95	.22
Positive hits	.57	.20
Aggregate model	.95	.47

high or low are shown in the right column. Although ΔR maintains a high correlation with ϕ (r = .97), the other strategies' correlations drop considerably. Note the dramatic decrease for the two strategies that use only Row 1 information, Proportion of Hits and Hits Minus False Positives.

The results regarding directional relation are shown in Table 4. The pattern and number of errors change with p(Outcome). As expected, most of the strategies often falsely imply a negative relation when p(Outcome) is low, often falsely imply a positive relation when p(Outcome) is high, and perform well when p(Outcome) is near .50.

In short, the preceding analyses show that (a) under some general conditions, the intuitive covariation strategies are quite redundant with each other and the normative strategy; (b) although suboptimal, ΔR performed virtually perfectly; (c) the strategies perform remarkably well when there is a reasonably strong relation to be detected; and (d) as p(Outcome) moves away from .5, the accuracy of most of the strategies decreases. In particular, as p(Outcome) decreases, the likelihood of falsely implying a negative relation increases, and as p(Outcome) increases, the likelihood of falsely implying a positive relation increases. Some basic knowledge regarding the environmental conditions under which a particular intuitive

TABLE 4
Percentage of Times Various Covariation Strategies Incorrectly Imply a Positive or Negative Relation for Different p(Outcome)

p(Outcome)	Positive relation error	Negative relation error
	Sum of diagonals	
$\leqslant$.1	13.1	12.4
$\geqslant$.45 and $\leqslant$.55	1.3	.7
$\geqslant$.9	13.1	12.4
	Positive testing	
$\leqslant$.1	.0	44.8
$\geqslant$.45 and $\leqslant$.55	9.3	6.4
$\geqslant$.9	44.8	.0
	Proportion of hits	
$\leqslant$.1	.0	49.5
$\geqslant$.45 and $\leqslant$.55	4.3	3.0
$\geqslant$.9	49.3	.0
	Aggregate model	
$\leqslant$.1	.2	40.3
$\geqslant$.45 and $\leqslant$.55	4.4	4.1
$\geqslant$.9	34.4	1.8

Note. ΔR is not shown because it is always correct, Positive Hits is not shown because there is no straightforward way to determine direction of relation, and Hits Minus False Positives is not shown because it is identical to Proportion of Hits.

strategy is used allows for predicting how likely an error is and what kind of an error it will be.

BAYESIAN INFERENCE

Because Bayes' theorem makes explicit use of prior probabilities, it can be used to study the interaction of prior beliefs with new information. People's intuitive ability to update their beliefs is of concern to areas of research such as hypothesis testing (Fischhoff & Beyth-Marom, 1983), categorization (Anderson, 1991a), memory (Anderson & Milson, 1989), attribution theory (Ajzen & Fishbein, 1975), jury decision making (Koehler & Shaviro, 1990), medical diagnosis (Casscells, Schoenberger, & Grayboys, 1978), and auditor judgment (Smith & Kida, 1991).

The following is Bayes' theorem in odds form:

$$p(H1)/p(H2) \times p(D|H1)/p(D|H2) = p(H1|D)/p(H2|D), \quad (1)$$

where $H1$ and $H2$ refer to two mutually exclusive hypotheses and D refers to data. Reading from the left, the first ratio is the prior odds, consisting of the probabilities of each hypothesis before the receipt of new information (D). The second ratio is the likelihood ratio, consisting of the probabilities of observing the new information, given the respective hypotheses are true. On the right is the posterior odds, which represent the normative outcome of the prior beliefs combined with new information.

In the empirical literature, one can distinguish between two types of Bayesian tasks. The first is what might be called a "one-shot" task, in which a subject is given base rates as a proxy for prior probabilities, sometimes given the two components of the likelihood ratio, and then asked to make a single response. In the second type of task, subjects see more than one datum and make more than one judgment for the same set of hypotheses. This type of task more closely corresponds to a "belief-updating" task, differing from the one-shot task not only because subjects make multiple judgments, but also because after the first response, subjects' prior probabilities rather than base rates are inserted into Bayes' theorem in order to determine the "correct" posterior probability.

In both kinds of tasks, subjects' responses often deviate from the Bayesian response. For one-shot tasks, a common finding is that subjects underweight or ignore the base rates (e.g., Bar-Hillel, 1980; Casscells et al., 1978; Doherty et al., 1979; Lyon & Slovic, 1976; Kahneman & Tversky, 1973). For belief-updating tasks, a common finding is that subjects average prior beliefs with new information (Hogarth & Einhorn, 1992; Lopes, 1985, 1987; Shanteau, 1970, 1972, 1975; Troutman & Shanteau, 1977; see also Anderson, 1981; Nisbett, Zukier, & Lemley, 1981; Slovic & Lichtenstein, 1971; Tetlock & Boettger, 1989).

The plan of this section is to discuss both kinds of Bayesian tasks, the empirical findings, how the normative and intuitive responses map onto a 2 × 2 matrix, and how subjects' apparent strategies for coping with Bayesian inference tasks compare with the normative strategy under various conditions.

One-Shot Tasks

A classic example of a Bayesian one-shot task is the Cab Problem (from Tversky & Kahneman, 1982a):

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (a) 85% of the cabs in the city are Green and 15% are Blue.
>
> (b) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.
>
> What is the probability that the cab involved in the accident was Blue rather than Green?

Bayes' theorem prescribes how to answer the question. $H1$ and $H2$ correspond to the hypothesis that the cab was Blue and the cab was Green, respectively, and D corresponds to the fact that the witness identified the cab as "Blue." The Bayesian response, then, is: $p(\text{Blue}|\text{``Blue''})/p(\text{Green}|\text{``Blue''}) = .80/.20 \times .15/.85 = .12/.17$. Because the hypotheses are exhaustive, the probability that the cab was Blue, given that the witness said "Blue," is $.12/(.12 + .17)$, which equals .41.

Bayesian problems such as these can be represented in 2 × 2 form. Let Rows 1 and 2 correspond to $H1$ and $H2$, respectively, and let Columns 1 and 2 correspond to Datum 1 and 2. Figure 2 illustrates how the Cab Problem maps onto a 2 × 2 matrix. The two hypotheses are that the cab was Blue or Green, and the data are that the witness says either "Blue" or "Green." The marginals for Rows 1 and 2 reflect the respective base rates for $H1$ and $H2$. Each cell in the matrix reflects how often the datum would be expected to occur in conjunction with the particular hypothesis. For example, because 15% of the cabs are Blue and the witness correctly identifies 80% of the Blue cabs, 12% of identifications would result in correctly identifying a Blue cab as "Blue" (Cell A). Similarly, because the witness incorrectly identifies Blue cabs as "Green" 20% of the time, 3% of identifications will fall into Cell B, and so on.

In this context, then, the prior odds are captured by $(A + B)/(C + D)$. The likelihood ratio corresponds to $[A/(A + B)]/[C/(C + D)]$. The posterior odds are A/C, and thus, the posterior probability of $H1$ is $A/(A + C)$, or $12/(12 + 17) = .41$.

Hypotheses	Data: "Blue"	Data: "Green"	
H1 = Blue Cab	Cell A 12	Cell B 3	15
H2 = Green Cab	Cell C 17	Cell D 68	85
			N = A+B+C+D

$$p(H1) = (A+B)/N$$

$$p(H2) = (C+D)/N$$

$$p(D|H1) = A/(A+B)$$

$$p(D|H2) = C/(C+D)$$

$$p(H1|D) = A/(A+C)$$

$$p(Data) = (A+C)/N$$

FIG. 2. The Cab Problem represented in 2 × 2 form.

How do people answer one-shot Bayesian problems? Subjects' median answer to the Cab Problem is .80, which corresponds to the witness' accuracy and apparently ignores the base rates of cabs in the city. Thus, one answer to the question is that people ignore base rates and use only the information in the likelihood ratio. (For other examples of ignoring base rates—or at least vastly underweighting them—see Bar-Hillel, 1980, 1990; Casscells et al., 1978; Doherty et al., 1979; Ginossar & Trope, 1980; Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982; Lyon & Slovic, 1976; Nisbett & Borgida, 1975.)

It is important to point out that subjects do not always ignore base rates in one-shot tasks. For example, it has been proposed that base rates are used if they are manipulated as a within-subjects variable (Birnbaum & Mellers, 1983; Fischhoff, Slovic & Lichtenstein, 1979; Kahneman & Tversky, 1982; Tversky & Kahneman, 1982), perceived as causally linked to the task at hand (Ajzen, 1977; Tversky & Kahneman, 1982), sufficiently specific (Bar-Hillel, 1980), or if the individuating data are not diagnostic (Fischhoff & Bar-Hillel, 1984; Ginossar & Trope, 1980; for a review of the base-rate literature, see Koehler, 1993). Although researchers have examined the conditions under which subjects do and do not use base rates in one-shot tasks, how the information is integrated into a final response is unclear. What may be the case, however, is that when base rates are used, they are sometimes "averaged" with the likelihood information. If so, this behavior would be similar to that found in belief-updating tasks described below.

Belief-Updating Tasks

As mentioned, many experimental Bayesian tasks involve updating probability estimates for the same hypothesis multiple times as new information is received. For example, imagine that the Cab Problem involved multiple witnesses, and after hearing whether each witness thought the cab was Blue or Green, you had to update the probability that the cab involved in the accident was in fact Blue rather than Green. If your response after the first witness was 80%, .80/.20 would then correspond to the prior odds. Given the second witness' testimony and accuracy, you would be asked for a second response. Normatively, this situation is exactly like the one discussed under one-shot tasks, except that the posterior odds given for response n become the prior odds for response $n + 1$. Thus, the prior odds are not given by the experimenter to the subject, but by the subject to the experimenter.

The phenomenon of interest here is the following: Assume that your response to the original Cab Problem is 80% and a second witness also identified the cab as "Blue," but this witness correctly identifies only 70% of both Blue and Green cabs. Because the witness identifies cabs at better than chance level, your posterior probability should increase to greater than 80% (the Bayesian response is 90%). Subjects, however, when confronted by strong support for a hypothesis followed by weak support for the same hypothesis, sometimes show a decrease in confidence, falling somewhere between the confidence levels that the subject would exhibit for each datum alone. When presented with the above two-witness scenario, 24 of 29 subjects responded with 75% (Bar-Hillel, 1980; see also Lopes, 1985, 1987; Nisbett et al., 1981; Shanteau, 1970, 1972, 1975; Troutman & Shanteau, 1977; see Hogarth & Einhorn, 1992, for an investigation of the conditions under which these errors occur). Such responses imply some sort of averaging of the prior and likelihood information (computationally speaking).

Following are the normative Bayesian strategy and four plausible strategies that either use only likelihood information, or average likelihood and base-rate information. (The strategies are defined below in traditional Bayesian notation. For the 2×2 matrix counterparts that were used in the simulation, see Fig. 2.).

Bayesian Inference Strategies

Bayes' theorem. For the purposes of the simulation, the normative response to the tasks discussed in this section is the Bayesian one, in which the probability that $H1$ is true, given the data, is $p(H1)p(D|H1)/[p(H1)p(D|H1) + p(H2)p(D|H2)]$.

Relative likelihood. As described in the section on one-shot Bayesian tasks, subjects sometimes ignore base rates. One possible strategy is to use the relative likelihood of $H1$, captured by $p(D|H1)/[p(D|H1) + p(D|H2)]$.[4]

Likelihood. Another possible base-rate neglect strategy is to use only the numerator of the likelihood ratio, where $H1$ is the hypothesis perceived to be of interest (so-called "pseudodiagnosticity"; see Beyth-Marom & Fischhoff, 1983; Doherty et al., 1979; Fischhoff & Beyth-Marom, 1983). In the Cab Problem, $H1$ = "the cab was Blue," and the likelihood information is that the witness correctly identified Blue cabs 80% of the time. The witness' accuracy for Green cabs is irrelevant here. Likelihood equals $p(D|H1)$.[5]

Relative likelihood average. One way to employ base rates is to average Relative Likelihood and the $H1$ base rate. This takes the form .5(Relative Likelihood) + $.5[p(H1)]$.[6]

Likelihood average. Another plausible strategy is to average Likelihood and the $H1$ base rate. For example, a subject might average the witness' accuracy in identifying Blue cabs (80) and the percentage of Blue cabs in the city (15). This is equal to $.5[p(D|H1)] + .5[p(H1)]$.

Method

The 6.25 million 2×2 matrices were generated based on all combinations of cell values between 1 and 50, and each strategy's output was calculated according to the above formulas. The product-moment correlation matrix was calculated, as well as whether or not the strategies were correct in terms of identifying $H1$ as more likely than $H2$, or vice versa.

Results and Discussion

Performance under the general conditions. Table 5 shows the product-moment correlation matrix for the five strategies based on the 6.25 million matrices. Most important is the first row, which shows the correlations between the normative strategy and each intuitive strategy. Note that Relative Likelihood Average correlates almost perfectly with the normative Bayesian response.

[4] Kahneman and Tversky's (1971; Tversky & Kahneman, 1982b) "representativeness" heuristic may be similar to this strategy (Gigerenzer & Murray, 1987).

[5] This is responding with $p(D|H1)$ instead of $p(H1|D)$. See Dawes (1988), Eddy (1982), Einhorn and Hogarth (1986), and Moskowitz and Sarin (1983) on confusing conditional probabilities.

[6] The averaging strategies use equal weighting for simplicity. Also, note that averaging would result in "conservative" judgments (Edwards, 1968), that is, judgments that are closer to .5, relative to the normative Bayesian response.

TABLE 5
Correlations between the Bayesian Inference Strategies' Outputs under the General Conditions

Strategy	2	3	4	5
1. Bayes' theorem	.79	.49	.99	.78
2. Relative likelihood	—	.67	.79	.60
3. Likelihood		—	.49	.83
4. Relative likelihood average			—	.79
5. Likelihood average				—

The middle column of Table 6 shows how often each strategy favors the correct hypothesis. The averaging strategies outperform the non-averaging strategies, and Relative Likelihood Average is always correct.[7]

Are the hypothesis-supporting errors "big" ones? As with the covariation strategies, one might wonder if many of the cases in which the intuitive Bayesian strategies support the wrong hypothesis are "small" errors in that the normative strategy weakly supports one hypothesis and the intuitive strategy weakly supports the other. In order to get a partial answer to this question, matrices for which the normative Bayesian response only weakly favored one hypothesis were eliminated. More specifically, the proportion of times each strategy supported the correct hypothesis was calculated for only those matrices for which the normative $H1$ posterior probability was either less than or equal to .4 or greater than or equal to .6 (4,165,000 matrices). The results are shown in the right column of Table 6 and indicate a substantial improvement in performance.

Environmental conditions that affect performance. Because it is well known that subjects sometimes ignore base rates (e.g., Bar-Hillel, 1980; Kahneman & Tversky, 1973), a starting point for investigating environmental factors is to examine the effects of different base rates on the strategies' performance. Relative Likelihood and Likelihood ignore base rates and would, therefore, be adversely affected by extreme base rates. However, because Relative Likelihood Average and Likelihood Average make use of base rates, these strategies should not be adversely affected.

Three subsets of the 6.25 million matrices were examined: Those matrices for which the $H1$ base rate (i.e., $[A + B]/N$) was (a) less than or equal to .1 (34,080 matrices), (b) between .45 and .55, inclusive (1,663,424 matrices), and (c) greater than or equal to .9 (34,080 matrices).

The correlations between the normative Bayesian response and the intuitive strategies when the base rate is near .5 are shown in the middle column of Table 7, and the correlations when the base rate is extreme are

[7] The results are virtually identical when cell values between 1 and 10 are used.

TABLE 6
Percent Correct of Accepting or Rejecting H1 for the Intuitive Bayesian Inference Strategies under Different Conditions

Intuitive strategy	All matrices	$p(H1 \mid D) \leq .4$ or $\geq .6$
Relative likelihood	75.0	83.3
Likelihood	66.7	72.2
Relative likelihood average	100.0	100.0
Likelihood average	79.6	88.8

shown in the right column. As expected, because Relative Likelihood and Likelihood ignore base rates, extreme base rates lead to a decrease in accuracy. In contrast, because Likelihood Average makes use of base rates (but ignores particular likelihood information), extreme base rates *aid* performance. Relative Likelihood Average is robust across conditions.

The results regarding how often the strategies incorrectly accept or reject *H*1 are shown in Table 8. (Relative Likelihood Average is not shown because it is always correct.) The pattern and number of errors change with the base rate. The two strategies that ignore base rates (Relative Likelihood and Likelihood) often incorrectly accept *H*1 when the *H*1 base rate is low and often incorrectly reject *H*1 when the base rate is high. Again, because Likelihood Average incorporates base rates, errors are fewest when the base rates are extreme.

There is another environmental factor to consider. Like many of the covariation strategies, two of the Bayesian strategies make more use of information in Cells *A* and *B* than Cells *C* and *D*. Likelihood uses only Cells *A* and *B* and, although Likelihood Average takes the *H*2 base rate into account, it ignores the *H*2 likelihood. Thus, analogous to the effects of *p*(Outcome) on the intuitive covariation strategies, differences in the probability of observing the data (i.e., *p*[Data]) may affect the intuitive Bayesian inference strategies. Using the Cab Problem as an example, the strategies may be affected if the witness often said "Blue" when asked to identify cabs, or often said "Green."

TABLE 7
Correlations between the Outputs of the Normative Bayesian Strategy with Each Intuitive Bayesian Strategy for Different Base Rates

	H1 Base rate	
Intuitive strategy	$\geq .45$ and $\leq .55$	$\leq .1$ or $\geq .9$
Relative likelihood	.99	.19
Likelihood	.66	.09
Relative likelihood average	.99	.98
Likelihood average	.68	.93

TABLE 8
Percentage of Times the Intuitive Bayesian Inference Strategies Incorrectly Accepted or Rejected H1 for Different Base Rates

H1 Base rate	Incorrectly accept H1	Incorrectly reject H1
	Relative likelihood	
≤.1	49.5	.0
≥.45 and ≤.55	3.0	4.3
≥.9	.0	49.3
	Likelihood	
≤.1	58.0	.2
≥.45 and ≤.55	12.5	13.1
≥.9	.0	48.5
	Likelihood average	
≤.1	.0	.9
≥.45 and ≤.55	12.1	12.5
≥.9	.0	.9

Note. Relative Likelihood Average is not shown because it is always correct.

Three subsets of the matrices were examined: Those matrices for which p(Data) (i.e., $[A + C]/N$) was (a) less than or equal to .1 (34,080 matrices), (b) between .45 and .55, inclusive (1,663,424 matrices), and (c) greater than or equal to .9 (34,080 matrices).

The correlations are shown in Table 9. Because Relative Likelihood takes into account both $p(D|H1)$ and $p(D|H2)$, p(Data) has minimal effects on the strategy. However, p(Data) has large effects on Likelihood and Likelihood Average, which ignore $p(D|H2)$; when p(Data) is extreme, performance decreases dramatically. Relative Likelihood Average is again robust across conditions.

The results regarding how often the intuitive strategies incorrectly accept or reject $H1$ are shown in Table 10. Again, the pattern and number of errors change with the environmental variable. The two strategies that

TABLE 9
Correlations between the Outputs of the Normative Bayesian Strategy with Each Intuitive Bayesian Strategy for Different p(Data)

	p(Data)	
Intuitive strategy	≥.45 and ≤.55	≤.1 or ≥.9
Relative likelihood	.76	.66
Likelihood	.66	.09
Relative likelihood average	.99	.99
Likelihood average	.96	.27

TABLE 10
Percentage of Times the Intuitive Bayesian Inference Strategies Incorrectly Accepted or Rejected H1 for Different *p*(Data)

p(Data)	Incorrectly accept H1	Incorrectly reject H1
	Relative likelihood	
≤.1	5.4	13.7
≥.45 and ≤.55	12.2	12.8
≥.9	15.0	15.8
	Likelihood	
≤.1	.2	58.0
≥.45 and ≤.55	13.1	12.5
≥.9	48.5	.0
	Likelihood average	
≤.1	.0	58.4
≥.45 and ≤.55	5.2	1.3
≥.9	44.5	.0

Note. Relative Likelihood Average is not shown because it is always correct.

ignore $p(D|H2)$ (Likelihood and Likelihood Average) often incorrectly reject $H1$ when *p*(Data) is low, often incorrectly accept $H1$ when *p*(Data) is high, and perform best when *p*(Data) is near .50.

In summary, the preceding analyses show that: (a) Relative Likelihood Average, which uses all the pertinent information but combines it inappropriately, performed virtually perfectly. (b) When one of the hypotheses was truly supported reasonably strongly, there were substantial increases in the strategies' performance. (c) Varying the base rates affects certain strategies in predictable ways. Strategies that ignore base rates lead to incorrectly accepting $H1$ when the base rate is low and incorrectly rejecting $H1$ when the base rate is high. (d) Varying *p*(Data) affects strategies that ignore $p(D|H2)$, the denominator in the likelihood ratio. When *p*(Data) is low, the strategies often result in incorrectly rejecting $H1$, and when *p*(Data) is high, they often result in incorrectly accepting $H1$. As with the covariation strategies, some basic knowledge regarding the environmental conditions under which a particular intuitive Bayesian strategy is used allows for predicting how likely an error is and what kind of an error it will be.

GENERAL DISCUSSION

In this section, I discuss three implications of the analysis of covariation assessment and Bayesian inference: (a) The accuracy of people's intuitive strategies, (b) a commonality between the intuitive strategies for the two tasks, and (c) how to improve accuracy.

The Accuracy of Intuitive Judgment Strategies

When examined under the general conditions of the simulations, all the apparent intuitive strategies performed much better than chance. However, this is not too surprising because each strategy makes use of at least some of the pertinent information. Indeed, the strategies' relative performance is largely due to the amount of information used. For example, covariation strategies that use more of the cells are generally more accurate, as are Bayesian strategies that do not ignore base rates. Nonetheless, many discussions regarding the validity of people's strategies have been discouraging (see, e.g., Nisbett & Ross, 1980, on people's ability to assess covariation between two dichotomous variables; Doherty et al., 1979, on ignoring the denominator of the likelihood ratio, $p(D|H2)$, in Bayesian inference). In this regard, perhaps the most interesting finding was that Relative Likelihood Average (i.e., averaging prior and likelihood information) in Bayesian inference results in virtually perfect performance according to the two accuracy measures used. This in spite of the fact that the strategy can lead to decreases in confidence when it should increase, and vice versa.[8] Thus, what may appear dysfunctional at a local level (i.e., case by case) may in fact be functional when a more global view is taken (i.e., across cases; see also Arkes, 1991; Brunswik, 1952; Campbell, 1959; Funder, 1987; Hogarth, 1981; Simon, 1955, 1956; Toda, 1962). Even strategies that resulted in considerably less than perfect accuracy may be reasonable to use. The utility of an intuitive strategy will be found in the combination of its cognitive simplicity and robustness. The simulation results under the general conditions suggest that, relative to the normative strategies, some intuitive strategies reduce accuracy only slightly, while considerably reducing cognitive load. Presumably, normative strategies are more taxing than intuitive strategies, and strategies that use more information are more taxing than those that use less. There is, however, no direct evidence of cognitive load (see Footnote 1).

[8] How can Relative Likelihood Average lead to decreases in confidence when it should increase (and vice versa), yet always support the correct hypothesis? Consider cases in which the prior information favors one hypothesis and the likelihood information favors the other. The normative Bayesian strategy supports the hypothesis that has the stronger support from either the prior or likelihood information. This is the same hypothesis that Relative Likelihood Average supports: The strategy averages prior and likelihood information. Clearer differences between the two strategies emerge when both the prior and likelihood information favor the same hypothesis. Here, the normative strategy leads to confidence more extreme than either the prior or the likelihood information (because Bayes' theorem is multiplicative), whereas the averaging strategy leads to confidence midway between the prior and likelihood information. Note, though, that even here the averaging strategy supports the correct hypothesis, just not strongly enough.

Thus, analyzing the effort/accuracy trade-off for judgment strategies is a potentially valuable avenue of research (for such an analysis of choice strategies, see Payne, Bettman, & Johnson, 1988, 1990; see also Beach & Mitchell, 1978).

A methodological implication of the redundancy between the intuitive and normative strategies is that correlation is a poor measure for determining which strategy a subject is using. In the area of covariation assessment, researchers have noted this (see Shaklee, 1983; Shaklee & Wasserman, 1986; Wasserman et al., 1990) and have contrived matrices more carefully in order to better distinguish between strategies. Another implication of the redundancy, however, is that the covariation task may be a prime candidate for invoking simple strategies because they may often result in only small decreases in accuracy.

The simulation results also show that most of the intuitive strategies' accuracy is strongly influenced by some simple environmental variables. For example, the performance of most of the strategies suffered when the Column 1 marginal was large or small relative to the Column 2 marginal, corresponding to differences in p(Outcome) and p(Data) for covariation assessment and Bayesian inference, respectively. Furthermore, the results show that most of the strategies make different *kinds* of errors under different conditions (i.e., incorrectly implying a positive or negative relation, or incorrectly accepting or rejecting $H1$). Thus, the cost/benefit trade-off of using a particular strategy depends not only on the accuracy versus simplicity of a strategy, but also on the relative costs of the two kinds of error (see, e.g., Friedrich, 1993).

Although we are presumably concerned with understanding how we perform in our natural environment, it is most often in the laboratory that judgment performance is examined. The laboratory environment is usually designed so that the experimenter can easily distinguish between an optimal model and the intuitive strategy (or strategies) of interest. Thus, performance based on intuitive strategies will often be poor in the laboratory. For example, if testing to see if subjects use the representativeness heuristic (Kahneman & Tversky, 1972), a researcher would present subjects with questions involving extreme base rates because it is under these conditions that the Bayesian response and the representativeness response are most easily distinguishable. The simulation results show that extreme base rates (the norm in the laboratory) lead to unusually poor performance for some of the intuitive Bayesian inference strategies. Thus, generalizing about the accuracy of intuitive strategies based on the laboratory environment may be misleading. Conclusions regarding the accuracy of intuitive judgment strategies appear incomplete without taking environmental conditions into account (see also Anderson, 1991b;

Brunswik, 1952, 1956; Campbell, 1959; Christensen-Szalanski & Beach, 1984; Funder, 1987; Hammond, 1955; Hammond, Hamm, Grassia, & Pearson, 1987; Hogarth, 1981; Simon, 1955, 1956, 1981; Toda, 1962).

The preceding discussion implicitly assumes that people's strategies are independent of the various environmental conditions under which people operate. It may be, however, that people's strategies reflect different environmental conditions in adaptive ways. The choice literature indicates that people invoke different choice strategies when normatively irrelevant task variables are changed (e.g., the number of alternatives; Einhorn & Hogarth, 1981; Payne, 1982). However, the changes in strategy reflect an adaptive effort/accuracy trade-off (Payne, Bettman, & Johnson, 1988, 1990). Analogous behavior may occur in judgment tasks. For instance, do people use different covariation strategies depending on p(Outcome)? One could use simple strategies and be highly accurate when p(Outcome) is near .50, but more sophisticated strategies are needed when p(Outcome) becomes extreme. Similarly, do differing concerns for incorrectly rejecting or accepting $H1$ affect intuitive Bayesian judgment strategies? If incorrectly rejecting $H1$ is of primary concern, then taking into account the $H1$ base rate becomes more important as it increases, and taking into account p(Data) becomes more important as it decreases. Subjects appear to have several strategies from which to choose. To what extent do environmental variables determine which strategy is invoked?

A Common Judgment Strategy

Mapping covariation assessment and Bayesian inference onto a 2 × 2 matrix allows for a simple comparison of subjects' strategies between the two tasks. In fact, the framework underscores an apparent theme, namely, that subjects often approach each task as a hypothesis-testing task and assess the degree to which the data are consistent with the one hypothesis perceived to be of interest.

In covariation assessment, it was argued that subjects often examine only Cells *A* and *B;* that is, they examine what happens in the presence of the Event. If subjects tend to approach the covariation task as a hypothesis-testing task and assess the likelihood of only a single hypothesis, the natural way to frame the task is in terms of "If Event, then Outcome," instead of, say, "If no Event, then Outcome." Thus, information about what happens when the Event does not take place (i.e., Cells *C* and *D*) is not considered to be of interest (see also Beyth-Marom, 1982).

That subjects are concerned with assessing a single hypothesis can be seen in the Bayesian inference literature as well. Recall that in the belief-updating tasks in which subjects repeatedly assess the likelihood of the same set of hypotheses based on new information, subjects' confidence sometimes moves in the wrong direction. For example, when presented with strong evidence followed by weak evidence for a hypothesis, subjects' confidence sometimes decreases after the second piece of evidence. Normatively, because both pieces of information favor the hypothesis, confidence should increase. However, if the subject is only assessing the degree to which the evidence is consistent with the hypothesis in question, weak evidence would reduce confidence. This behavior is consistent with the two averaging models examined in the simulation. The analysis presented here, however, leads to the hypothesis that Likelihood Average is the more common strategy because it ignores the impact of the data on *H*2.

A similar prediction can be made regarding one-shot Bayesian tasks. Recall that in the Cab Problem, because the accuracy of the witness is 80% for both Blue and Green cabs, it is unclear whether the median response of 80% is due to a consideration of both likelihoods (i.e., a comparison of accuracy rates for Blue versus Green cabs) or just due to the likelihood of *H*1 (i.e., the fact that the witness is correct 80% of the time for Blue cabs). The latter explanation is probably the best one. A simple test could be conducted through orthogonally varying the witness' accuracy for Blue and Green cabs.

Klayman and Ha (1987) claim that subjects often test hypotheses according to a "positive hypothesis testing strategy" that involves checking for expected features (see also Fischhoff & Beyth-Marom, 1983). For example, if testing to see if a person is extraverted, subjects are most likely to look for features consistent with the extravert hypothesis, whereas introvert features are largely ignored. Again, the strength of the alternative hypothesis (i.e., How many introvert features are present?) is not taken into account. Thus, Klayman and Ha's thesis regarding how people test hypotheses is consistent with the present account of covariation assessment and Bayesian inference: A single hypothesis is perceived to be of interest, and the subjective probability that the hypothesis is true is a function of the degree to which the data are consistent with it. Put more abstractly, subjects are assessing the degree to which "If *X*, then *Y*," where *X* is the hypothesis of interest, and *Y* is the expected data. Of interest is that, when asked directly to test (deterministic) statements of the form "If *X*, then *Y*," subjects consider irrelevant what happens when not-*X* is the case (Evans, 1972; Johnson-Laird & Tagart, 1969; see also Evans, 1982, 1989; Johnson-Laird and Byrne, 1991; Wason, 1966; Wason & Johnson-Laird, 1972).

Improving Accuracy

If subjects tend to assess the one hypothesis perceived to be of interest in covariation assessment and Bayesian inference, then simply considering the alternative hypothesis is crucial to better performance. However, this does not necessarily imply using a normative strategy (i.e., calculating the ϕ coefficient or using Bayes' theorem). Some of the intuitive strategies take into account the alternative hypothesis and are accurate across environments.

There is evidence for both tasks that subjects' performance improves when they are encouraged to *compare* hypotheses. In the case of covariation assessment, Shaklee and Mims (1981) found that 76% of their adult subjects appeared to use strategies involving all four cells (i.e., ΔR or Sum of Diagonals) when asked to assess whether the Outcome was more or less likely in the Event's presence than in the Event's absence, a phrasing of the task that encourages a comparison of the two hypotheses.[9] Considering the alternative hypothesis when assessing covariation (i.e., "Does the absence of the Event lead to the Outcome?") might result in the use of ΔR, which the simulation results show is virtually perfectly correlated with the ϕ coefficient.

In a Bayesian inference task, Beyth-Marom and Fischhoff (1983, Experiment 1) found that when phrasing the question in comparative terms (i.e., "Is $H1$ or $H2$ more probable?"), 78% of the subjects claimed that $p(D|H2)$ was important to know, almost equaling the 83.1% who were interested in $p(D|H1)$. When phrased as "How probable is $H1$?," only 53.6% of the subjects wanted to know $p(D|H2)$. For belief updating, subjects might still use an averaging strategy, but when $H2$ is considered, the simulation results show that the strategy (Relative Likelihood Average) correlates virtually perfectly with the normative Bayesian response.

Taken together, these experiments support the contention that subjects usually cope with covariation assessment and Bayesian inference by evaluating a single hypothesis, but that they compare two hypotheses when that is the perceived task (see also Bassok & Trope, 1984; Gorman, Stafford, & Gorman, 1987; Skov & Sherman, 1986; Trope & Bassok, 1982, 1983; Trope, Bassok, & Alon, 1984; Tweney et al., 1980). Although considering the alternative hypothesis does not entail using a normative strategy, it can lead to using intuitive strategies that appear relatively simple and are accurate across environments. Using these strategies may be the most efficient means of ensuring highly accurate judgment in these important tasks.

[9] Although it is difficult to make direct comparisons between the Shaklee and Mims (1981) study and studies not using the comparative phrasing, studies typically find fewer subjects using all four cells. For example, Shaklee and Tucker (1980) found that 58% of their subjects used strategies involving all four cells, and Wasserman et al. (1990) found 33% and 34% of their subjects using all four cells in Experiments 2 and 3, respectively.

SUMMARY

Although it has long been recognized that people often do not process information in the manner prescribed by normative models of judgment, it has been unclear how often and to what degree one should expect to be led astray by using intuitive strategies. Through Monte Carlo simulation, the accuracy of people's strategies for coping with covariation assessment and Bayesian inference was examined. Under some general conditions, all the strategies performed better than chance and many performed surprisingly well. Furthermore, some simple environmental variables were found to have large effects on the accuracy of most of the intuitive strategies. Indeed, some basic knowledge regarding the environmental conditions under which a particular intuitive strategy is used allows for predicting how likely an error is and what kind of an error it will be (e.g., incorrectly rejecting versus incorrectly accepting $H1$). Conclusions regarding the accuracy of intuitive judgment strategies appear incomplete without taking into account the structure of the environment in which a particular strategy is used.

It was also argued that subjects approach both covariation assessment and Bayesian inference in essentially the same manner. In particular, subjects often appear to assess the degree to which the data are consistent with the hypothesis of interest and do not take into account the degree to which the data are consistent with the alternative hypothesis. A key to better performance, then, lies in taking into account the strength of the alternative hypothesis, although this does not necessarily imply using a normative strategy (i.e., calculating the ϕ coefficient or using Bayes' theorem). *Intuitive* strategies that take into account the alternative hypothesis are accurate *across environments*. Because intuitive strategies are presumably simpler than normative strategies and are already part of people's repertoire, simply taking into account the alternative hypothesis may be the most efficient means of accurately assessing covariation and updating beliefs.

REFERENCES

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, **35**, 303–314.

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, **82**, 261–277.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, **15**, 147–149.

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response. *Canadian Journal of Psychology*, **34**, 1–11.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgments of influence. *Learning and Motivation*, **14**, 381–405.

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

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, **91**, 112–149.

Anderson, J. R. (1991a). The adaptive nature of human categorization. *Psychological Review*, **98**, 409–429.

Anderson, J. R. (1991b). Is human cognition adaptive? *Behavioral and Brain Sciences*, **14**, 471–517.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, **96**, 703–719.

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, **110**, 486–498.

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, **112**, 117–135.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, **44**, 211–233.

Bar-Hillel, M. (1990). Back to base rates. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 200–216). Chicago: University of Chicago Press.

Bassok, M., & Trope, Y. (1984). People's strategies for testing hypotheses about another's personality: Confirmatory or diagnostic? *Social Cognition*, **2**, 199–216.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, **3**, 439–449.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, **10**, 511–519.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, **45**, 1185–1195.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, **45**, 792–804.

Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *The handbook of social psychology* (Vol. 2). Cambridge, MA: Addison-Wesley.

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.

Campbell, D. T. (1959). Systematic error on the part of human links in communication systems. *Information and Control*, **1**, 334–369.

Casscells, W., Schoenberger, A., & Grayboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, **299**, 999–1001.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, **72**, 193–204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, **74**, 271–280.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist*, **39**, 75–78.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, **90**, 272–292.

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, **8**, 214–220.

Dawes, R. M. (1988). *Rational choice in an uncertain world*. San Diego: Harcourt Brace Jovanovich.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, **43**, 111–121.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge: Cambridge University Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, **32**, 53–88.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, **99**, 3–19.

Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *The Quarterly Journal of Experimental Psychology*, **24**, 193–199.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. London: Erlbaum.

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory and Cognition*, **12**, 402–410.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, **90**, 239–260.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, **23**, 339–359.

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, **100**, 298–319.

Funder, D. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, **101**, 75–90.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Ginossar, Z., & Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology*, **16**, 228–242.

Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology*, **39A**, 1–28.

Hamilton, D. L. (1976). Cognitive biases in the perception of social groups. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 81–93). Hillsdale, NJ: Erlbaum.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, **12**, 392–407.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, **62**, 255–262.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, **17**, 753–771.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, **47**, 116–131.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, **90**, 197–217.

Hogarth, R. M. (1987). *Judgement and choice: The psychology of decision* (2nd ed.). Chichester, England: Wiley.

Hogarth, R. M. (Ed.). (1990). *Insights in decision making: A tribute to Hillel J. Einhorn.* Chicago: University of Chicago Press.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, **24**, 1–55.

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

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, **79**(1, Whole No. 594).

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211–230). Cambridge: Cambridge University Press.

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science*, **31**, 395–414.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* London: Erlbaum.

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology*, **82**, 367–373.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237–251.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science*, **12**, 313–327.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, **11**, 123–141.

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

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211–228.

Koehler, J. J. (1993). *The base rate fallacy reconsidered: Normative, descriptive, and methodological challenges.* Unpublished manuscript, University of Texas at Austin.

Koehler, J. J., & Shaviro, D. N. (1990). Veridical verdicts: Increasing verdict accuracy through the use of overtly probabilistic evidence and methods. *Cornell Law Review*, **75**, 247–279.

Lipe, M. G. (1990). A lens model analysis of covariation research. *Journal of Behavioral Decision Making*, **3**, 47–59.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, **39**, 821–831.

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Sex stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, **18**, 23–42.

Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society*, **23**, 509–512.

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica*, **64**, 167–185.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, **40**, 287–298.

McKenzie, C. R. M. (1993). *The perceived informativeness of covariation data: Causal direction and expected benefit*. Unpublished manuscript, University of Chicago.

Moskowitz, H., & Sarin, R. K. (1983). Improving the consistency of conditional probability assessments for forecasting and decision making. *Management Science*, **29**, 735–749.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, **29**, 85–95.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology*, **30**, 395–406.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, **32**, 932–943.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, **13**, 248–277.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, **92**, 382–402.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). The adaptive decision maker: Effort and accuracy in choice. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 129–153). Chicago: University of Chicago Press.

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 676–696.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, **110**, 101–120.

Seggie, J. L., & Endersby, H. (1972). The empirical implications of Piaget's concept of correlation. *Australian Journal of Psychology*, **24**, 3–8.

Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: Freeman.

Shaklee, H. (1983). Human covariation judgment: Accuracy and strategy. *Learning and Motivation*, **14**, 433–448.

Shaklee, H., & Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Development*, **52**, 317–325.

Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 208–224.

Shaklee, H., & Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory & Cognition*, **8**, 459–467.

Shaklee, H., & Wasserman, E. A. (1986). Judging interevent contingencies: Being right for the wrong reasons. *Bulletin of the Psychonomic Society*, **24**, 91–94.

Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, **85**, 181–191.

Shanteau, J. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, **93**, 63–68.

Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica*, **39**, 83–89.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, **69**, 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, **63**, 129–138.

Simon, H. A. (1981). *The sciences of the artificial* (2nd ed.). Cambridge, MA: MIT Press.

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology*, **22**, 93–121.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, **28**, 1–39.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, **6**, 649–744.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, **4**, 165–173.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, J. F. & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin*, **109**, 472–489.

Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology*, **57**, 388–398.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, **25**, 219–225.

Toda, M. (1962). The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behavioral Science*, **7**, 164–183.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology*, **43**, 22–34.

Trope, Y., & Bassok, M. (1983). Information-gathering strategies in hypothesis testing. *Journal of Experimental Social Psychology*, **19**, 560–576.

Trope, Y., Bassok, M., & Alon, E. (1983). The questions lay interviewers ask. *Journal of Personality*, **52**, 90–106.

Troutman, C. M., & Shanteau, J. (1977). Inferences based on nondiagnostic information. *Organizational Behavior and Human Performance*, **19**, 43–55.

Tversky, A., & Kahneman, D. (1971). The belief in the "law of small numbers." *Psychological Bulletin*, **76**, 105–110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

Tversky, A., & Kahneman, D. (1982a). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). Cambridge: Cambridge University Press.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross,

K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *The Quarterly Journal of Experimental Psychology*, **32**, 109–123.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, **19**, 231–241.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, **12**, 129–140.

Wason, P. C. (1966). Reasoning. In M. B. Foss (Ed.), *New horizons in psychology* (pp. 135–161). Harmondsworth, England: Penguin.

Wason, P. C. (1968). "On the failure to eliminate hypotheses . . ."—A second look. In P. C. Wason & P. N. Johnson-Laird (Eds.), *Thinking and reasoning* (pp. 165–174). Harmondsworth, England: Penguin.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 509–521.

Wasserman, E. A., & Shaklee, H. (1984). Judging response-outcome relations: The role of response-outcome contingency, outcome probability, and method of information presentation. *Memory and Cognition*, **12**, 270–286.

(Accepted April 29, 1993)