SRRI #158

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES **40**, 255–269 (1987)

Understanding Conditional Probabilities

ALEXANDER POLLATSEK, ARNOLD D. WELL, CLIFFORD KONOLD, AND PAMELA HARDIMAN

University of Massachusetts, Amherst

AND

GEORGE COBB

Mount Holyoke College

In two experiments, subjects were asked to judge whether the probability of A given B was greater than, equal to, or less than the probability of B given A for various events A and B. In addition, in Experiment 2, subjects were asked to estimate the conditional probabilities and also to calculate conditional probabilities from contingency data. For problems in which one conditional probability was objectively larger than the other, performance ranged from about 25–80% correct, depending on the nature of A and B. Changes in the wording of problems also affected performance, although less dramatically. Patterns of responses consistent with the existence of a causal bias in judging probabilities were observed with one of the wordings used but not with the other. Several features of the data suggest that a major source of error was the confusion between conditional and joint probabilities.

Teachers of statistics generally agree that people have a great deal of difficulty with conditional probabilities. Some problems seem to be caused by formal wording or algebraic notation, although there is a surprising amount of difficulty even when the conditional probability is expressed in terms of simple percentages, such as "the percentage of smokers who get lung cancer."

Reports in the literature suggest that difficulties occur even with intelligent and highly educated people. For example, Eddy (1982) reported a tendency among physicians to confuse the predictive accuracy of an X-ray report with its retrospective accuracy, i.e., a confusion between p(cancer|positive test) and p(positive test|cancer). There is, however, little systematic data available on how well people understand different

The research described in this paper was supported by NSF Grants SED-8113323 and BNS85-09991. The authors thank Baruch Fischhoff for comments on an earlier draft. Requests for reprints should be sent to either Alexander Pollatsek or Arnold Well, Department of Psychology, University of Massachusetts, Amherst, MA 01003.

kinds of conditional probability statements and the kinds of difficulties they have with them.

There are several possible sources of error. People may have difficulty with the syntax of conditional probability statements so that performance depends on the details of the wording. A number of confusions are possible: the probability of A given B could be confused with the probability of B given A, the joint probability of A and B, or even the joint frequency of A and B. Einhorn and Hogarth (1986) found that statements employing the conjunction *and* could be interpreted as referring to either joint or conditional probabilities. Although a statistician would interpret "What is the probability of going into the supermarket and buying some coffee?" as a question about the joint probability p(going into the supermarket ∩ buying coffee), when the statement was presented to 24 graduate students who had taken at least one statistics course, 9 of them interpreted it as referring to p(buying coffee|going into the supermarket). Conversely, it is possible that an expression such as "the percentage of brown-haired men who have green eyes" which statisticians would interpret as a conditional probability might be interpreted by novices as a joint probability, especially if parsed as (the percentage of) (brown-haired men who have green eyes).

A second possible source of error is interference from causal reasoning. People may confuse conditionality with causality both because conditional probabilities are used in inferring causal relationships and because statements about both conditional probability and causality employ key words like *if* or *given that*. Moreover, even in cases in which there is only one causal relationship (e.g., A causes B but not the reverse), there will be two conditional probabilities, $p(A|B)$ and $p(B|A)$. Both conditional probabilities may be viewed as representing the same causal relationship and hence thought of as equal.

Errors may also follow from the belief that causal relationships should be stronger than diagnostic ones (i.e., p(effect|cause) > p(cause|effect)). Of course, p(effect|cause) need not be greater than p(cause|effect) in general. There are many situations in which an effect (e.g., the presence of an infectious disease) occurs only if multiple causes (e.g., disease bacteria, lack of antibodies) all occur. In this example, p(bacteria|disease) = 1 while p(disease|bacteria) < 1.

Tversky and Kahneman (1980) have argued that a causal bias exists in judging conditional probabilities and presented evidence in support of this position. For example, they asked subjects if it was more probable "(a) That a girl has blue eyes, if her mother has blue eyes; (b) That the mother has blue eyes, if her daughter has blue eyes; or (c) The two events are equally probable." There was a strong tendency to answer (a) instead of (b) (69 as opposed to 21 respondents), although almost half (75 respondents) chose the correct answer (c). However, it is difficult to interpret this finding as strong evidence for a "causal bias" in making judgments about conditional probabilities because subjects may have given their answers for several reasons, including not fully understanding the wording of the problem. A particular difficulty with the wording of this example is that it can be interpreted as requiring a judgment about causality rather than one of conditionality.

The present paper addresses several issues: First, do naive people have a fundamental inability to deal with conditional probabilities or do they have the necessary conceptual structure but are subject to confusions elicited by the wording of problems or the situations to which they refer? Second, when difficulty is encountered with a statement about $p(A|B)$, is there a tendency to confuse it with $p(B|A)$ or p(A and B)? Finally, is the kind of causal bias refered to by Tversky and Kahneman (1980) powerful and pervasive?

In the first experiment, subjects were given a series of problems in which they were asked which, if either, of two conditional probabilities was larger. In a second experiment, additional problems were presented, worded in terms of probabilities for half the subjects and in terms of percentages for the other half. Subjects were subsequently required to estimate both conditional probabilities and, in a final condition, to calculate two conditional probabilities from data that were presented.

EXPERIMENT 1

Method

Subjects. Eighty-six undergraduates taking lower division psychology courses at the University of Massachusetts who had not taken a college level statistics course were asked to complete a short questionnnaire on the first day of class. It was emphasized that participation was voluntary and that performance on the questionnaire would not influence course grades. Subjects were given 10 min to complete their answers.

Materials. The questionnaire consisted of six questions (see Table 1). For each, subjects were asked which of three options stated the correct relationship between $p(A|B)$ and $p(B|A)$: $p(A|B) < p(B|A)$, $p(B|A) > p(A|B)$, or $p(A|B) = p(B|A)$.

Three of the problems were chosen to pit implication against causality. In two cases, event A could be thought of as (1) a necessary but not sufficient cause of event B and (2) strongly implied by B. For example, in the sickness–fever problem, the event "being sick" (A) both causes and is implied by "having a fever" (B). In the second problem, A was an unhappy marriage and B was a divorce. In a third problem, A (marijuana use) is not strictly a cause of B (heroin use) nor does heroin use neces-

TABLE 1
METHODS AND RESULTS OF EXPERIMENT 1

	Percentage of responses
1. Which of the following events is more probable?	
(a) That a girl has blue eyes if her mother has blue eyes	35
(b) That a mother has blue eyes if her daughter has blue eyes	14
(c) The two events are equally probable (N = 86)	51
2. In which prediction would you have the greatest confidence?	
(a) Predicting that a person who has a fever is sick	87
(b) Predicting that a person who is sick has a fever	3
(c) Equal confidence in both predictions (N = 86)	10
3. In which of the following statements do you have the most confidence?	
(a) An unhappy marriage will result in divorce	8
(b) A divorce was the result of an unhappy marriage	72
(c) Equal confidence in both statements (N = 86)	20
4. Which of the following events is more probable?	
(a) That a person addicted to heroin also smokes marijuana	72
(b) That a person who smokes marijuana is also addicted to heroin	0
(c) The two events are equally probable (N = 81)	28
5. A cancer test was given to all residents of a large city. A positive test was indicative of cancer, and a negative test of no cancer. In which prediction do you have the greatest confidence?	
(a) Predicting that a person had cancer if they got a positive test	15
(b) Predicting a positive test for a person who had cancer	37
(c) Equal confidence in both predictions (N = 79)	48
6. Which of the following events is more probable?	
(a) That a blue cab is correctly identified at night as a blue cab	24
(b) That a cab identified at night as a blue cab is really blue	7
(c) The two events are equally probable (N = 85)	69

The correct answers are underlined. There are no correct answers to the last two problems unless additional information is provided.

sarily imply use of marijuana. However, because many people think that marijuana usage leads to heroin usage (which is close to a causal notion) and that virtually all heroin users also use marijuana, it seems to be a similar problem.

The other three problems were chosen to be sensitive to any causal bias that might have existed. One was the mother–daughter problem cited earlier. The other two were adpated from the cab and disease problems that have frequently been used in studies investigating the use of base-rate information (e.g., Bar-Hillel, 1980; Casscells, Schoenberger, & Graboys, 1978; Hammerton, 1973; Tversky & Kahneman, 1980). The cab and the cancer problems did not present information about the base rates of blue cabs and cancer or information about the accuracy of the witness or diagnostic test, so in neither case was there an objectively correct answer.

Results and Discussion

The percentage of subjects who chose each alternative are given in Table 1. The patterns of answers were quite different for the "implication" and the "causal bias" problems. Most of the subjects gave the correct answer for the former: 87, 72, and 72% for the fever–sickness, unhappy marriage–divorce, and marijuana–heroin problems, respectively. In all three cases, subjects strongly preferred the alternative consistent with $p(\text{cause}|\text{effect}) > p(\text{effect}|\text{cause})$. For the "causal bias" problems, answer (c), $p(A|B) = p(B|A)$, was the modal answer, being given by 51, 48, and 69% of the subjects for the mother–daughter, cancer, and cab problems, respectively. In each of these problems, however, more subjects chose the alternative that could be interpreted as indicating $p(\text{effect}|\text{cause}) > p(\text{cause}|\text{effect})$ than chose the remaining alternative, $\chi^2(1) = 11.52$, 7.05, and 7.54 for the mother–daughter, cancer, and cab problems, all $ps < .01$.

These results suggest that naive subjects are not invariably confused by statements dealing with conditional probabilities. For the three "implication" problems, about 77% of the responses were correct and the most common error was to judge that $p(A|B)$ and $p(B|A)$ were equal. Thus any causal bias that might have existed for these problems played at most a minor role. The remaining three problems were selected to be sensitive to causal bias. For these problems, there was a tendency to choose the alternative consistent with $p(\text{effect}|\text{cause}) > p(\text{cause}|\text{effect})$ as opposed to the reverse. However, the size of the effect was quite modest.

EXPERIMENT 2

The good performance on the three implication problems suggests that naive subjects can successfully deal with conditional probabilities, at least under some conditions. However, given the atypical nature of the conditionals employed in these problems, (i.e., $p(A|B) \approx 1$ and $p(B|A) << 1$), it seemed worthwhile to explore performance with additional problems for which neither of the conditional probabilities have values approaching 1.

The additional problems (see Table 2) were chosen so that it should have been possible to decide which conditional probability was larger on the basis of common real-world knowledge. For example, for the *teacher* problem, $p(\text{woman}|\text{school teacher})$ is at least .5, while $p(\text{teacher}|\text{woman})$ is much less than .5. In the *green-eyed man* problem, p(man|green-eyed

TABLE 2
MATERIALS USED IN EXPERIMENT 2

For each of the following cases, indicate which of the two events is more probable (which of the two percentages is bigger). If you think that the events are equally probable (the percentages are equal), indicate by circling (c). In each case try to comment on *why* you chose the answer you did.

Probability Version	Percentage Version
1. (a) that a person who has a fever is sick	1. (a) the percentage of people who have fevers that are sick
(b) that a person who is sick has a fever	(b) the percentage of sick people who have fevers
(c) the two are equally probable	(c) the two percentages are equal
2. (a) that a person who smokes marijuana is a heroin addict	2. (a) the percentage of people who smoke marijuana that are heroin addicts
(b) that a heroin addict smokes marijuana	(b) the percentage of heroin addicts that smoke marijuana
(c) the two are equally probable	(c) the two percentages are equal
3. (a) that a woman is a school teacher	3. (a) the percentage of women that are school teachers
(b) that a school teacher is a woman	(b) the percentage of school teachers that are women
(c) the two are equally probable	(c) the two percentages are equal
4. If we make a distinction between trucks and passenger vehicles	4. If we make a distinction between trucks and passenger vehicles
(a) that a Datsun is a passenger vehicle	(a) the percentage of Datsuns that are passenger vehicles
(b) that a passenger vehicle is a Datsun	(b) the percentage of passenger vehicles that are Datsuns
(c) the two are equally probable	(c) the two percentages are equal
5. (a) that a man has green eyes	5. (a) the percentage of men that have green eyes
(b) that a green-eyed adult is a man	(b) the percentage of green-eyed adults that are male
(c) the two are equally probable	(c) the two percentages are equal
6. (a) that a green-eyed person has brown hair	6. (a) the percentage of green-eyed people that have brown hair
(b) that a brown-haired person has green eyes	(b) the percentage of brown-haired people that have green eyes
(c) the two are equally probable	(c) the two percentages are equal
7. For families in which there is one daughter:	7. For families in which there is one daughter:
(a) that a girl has blue eyes if her mother has blue eyes	(a) the percentage of blue-eyed mothers who have blue-eyed daughters
(b) that a mother has blue eyes if her daughter has blue eyes	(b) the percentage of blue-eyed girls who have blue-eyed mothers
(c) the two are equally probable	(c) the two percentages are equal

adult) is about .5 and p(green eyes|man) is less than .5. In these additional "real-world knowledge" problems, neither event was likely to be perceived as a cause of the other, so that if subjects had difficulty with them, it would be unlikely to be due to causal reasoning strategies.

In addition, we manipulated the wording of the problems to get at least some idea of how much the pattern of responses depends on the exact wording used to express the conditional probabilities. Although both the *percentage version* (e.g., "the percentage of women who are teachers") and the *probability version* (e.g., "the probability that a woman is a teacher") referred to exactly the same concept, the former seemed to us to emphasize the empirical relative frequency, while the latter could possibly invoke the notion of prediction. It seemed especially important to compare performance on the two versions of the mother–daughter problem. Although performance on this problem in Experiment 1 was consistent with what would be expected if there was a causal bias in judgments about conditional probabilities, we really cannot rule out the possibility that the pattern of responses was elicited by the specific wording of the problem.

We also attempted to get some insights into subjects' reasoning by employing several additional tasks. A second section was added to the questionnaire which requested subjects to provide numerical estimates of the conditional probabilities. The estimation data could potentially help us understand what confusions led to the choice of incorrect alternatives, and the more demanding nature of the task might result in increased involvement and better performance. We also asked subjects to provide written explanations of their answers. Finally, at the end of the session, half the subjects received a small set of data from which they were instructed to calculate $p(A|B)$ and $p(B|A)$. If the correct conditional was not computed, the data could provide information on whether the major source of confusion was between the two conditionals or between the conditional and the joint probability.

Method

Subjects and procedure. One hundred twenty subjects were recruited from sections of an introductory psychology course designed for majors and received course credit for participation. None had previously taken a college statistics course.

Subjects were run in groups of 5–15 in 40-min sessions. To discourage quick answers, subjects were given fixed amounts of time to complete each section of the questionnaire and were not allowed to begin the next section until the time allotted to the current section had elapsed.

Materials. Two sections were given to all subjects. The first consisted of the seven problems presented in Table 2. There were two implication problems (*fever* and *marijuana*), one "causal bias" problem (*mother–daughter*), and four "real-world knowledge" problems that fit into neither of these categories. As in Experiment 1, subjects were asked to judge whether $p(A|B)$ was greater, less than, or equal to $p(B|A)$. Half the subjects received the probability version of the problems and the other half received the percentage version. The second section contained the same problems in the same order and version, but here the task was to estimate the conditional probabilities. Subjects were asked to give their estimates in percentages, as we had previously found subjects to be more comfortable with percentages than proportions. Subjects were again requested to justify their answers.

Sixty subjects received a third section in which they were provided with a table containing information about the eye color and hair color of 25 individuals. Each row in the table contained the initials of a hypothetical individual followed by information about the person's hair and eye color. Subjects were asked to calculate (1) "the percentage of green-eyed people that had brown hair" and (2) "the percentage of brown-haired people that had green eyes." For half the subjects, eye color was the first attribute listed in the table and for the other half it was hair color.

Results and Discussion

Forced-choice data. Two booklets were dropped from the analysis because they were almost completely blank. For the remaining 118 booklets, there were no more than four missing responses for any of the seven problems.

The forced-choice data presented in Table 3 indicate that performance varied widely across problems. More than 80% of the responses on the marijuana problem were correct but less than 30% were correct for the two "attribute" problems (eye–hair and green-eyed man). It is of interest that performance was quite good on the teacher and vehicle problems, suggesting that high levels of success are not restricted to problems for which one of the conditional probabilities is virtually equal to one.

Averaged across all seven problems, performance was similar for both wordings. The average rate of correct responses was 57.0% for the probability version and 56.7% for the percentage version. However, looked at problem by problem, the patterns of responses for the two versions differed significantly for four of the problems. For the mother–daughter problem, $\chi^2(2) = 7.94$, $p < .02$, the tendency to choose response *a* over response *b*, that had previously been interpreted as support for causal bias, did not occur in the percentage version. It appears as though something about the probability version elicited causal reasoning but that this rarely happened in the percentage version. It is also possible that subjects may have interpreted the wording of the probability version (but not that

TABLE 3

COMPARISON OF RESULTS FOR THE PERCENTAGE AND PROBABILITY VERSIONS OF THE PROBLEMS IN EXPERIMENT 2

Problem	Response	Probability version	Percentage version	Combined data	Number of respondents
	a	48 (83%)	36 (62%)	84 (72%)	116
Fever	b	2 (3%)	12 (21%)	14 (12%)	
	c	8 (14%)	10 (17%)	18 (16%)	
	a	0 (0%)	4 (7%)	4 (3%)	116
Marijuana	b	46 (79%)	48 (83%)	84 (81%)	
	c	12 (21%)	6 (10%)	18 (16%)	
	a	7 (12%)	3 (5%)	10 (9%)	116
Teacher	b	34 (59%)	38 (66%)	72 (62%)	
	c	17 (29%)	17 (29%)	34 (29%)	
	a	40 (69%)	38 (66%)	78 (67%)	116
Vehicle	b	1 (2%)	8 (14%)	9 (8%)	
	c	17 (29%)	12 (21%)	29 (25%)	
	a	20 (34%)	13 (22%)	33 (28%)	116
Green-eyed man	b	18 (31%)	15 (26%)	33 (28%)	
	c	20 (34%)	30 (57%)	50 (43%)	
	a	13 (23%)	14 (24%)	27 (24%)	114
Eye–hair	b	1 (2%)	10 (17%)	11 (10%)	
	c	42 (75%)	34 (59%)	76 (66%)	
	a	21 (35%)	8 (14%)	29 (25%)	115
Mother–daughter	b	5 (9%)	9 (16%)	14 (12%)	
	c	32 (55%)	40 (70%)	72 (63%)	

of the percentage version) as requiring a judgment about causality rather than about conditional probability. The results do not support the notion of a strong causal bias in judgments of conditional probabilities per se.

There were also significantly different patterns of responses for the two wordings of the fever problem, $\chi^2(2) = 9.07$, $p < .01$, the vehicle problem, $\chi^2(2) = 6.36$, $p < .05$, and the eye–hair problem, $\chi^2(2) = 8.23$, $p < .02$. In each case there were fewer reversals (i.e., fewer choices of $p(A|B)$ as larger when $p(B|A)$ should have been chosen) in the probability version. The number of reversals observed for the probability and percentage versions were 2 vs 12, 1 vs 8, and 1 vs 10 for the fever, vehicle, and eye–hair problems, respectively.

Estimation data. Subjects' estimates were first converted to forced-choice data by noting whether the estimate of $p(A|B)$ was less than greater than, or equal to the estimate of $p(B|A)$. To simplify discussion, we combine the data from the probability and percentage version of each problem. As can be seen in Table 4, there was almost 80% agreement between the forced-choice responses directly obtained in the first section of the questionnaire and those derived from the estimation data. How-

TABLE 4
RELATIONSHIP BETWEEN FORCED-CHOICE AND ESTIMATION RESPONSES

Problem	Forced-choice response	Choice from estimates: a	b	c	Total
	a	76	3	1	80
Fever	b	4	9	0	13
	c	10	1	8	19
	Total	90	13	9	112
	a	1	2	1	4
Marijuana	b	2	87	2	91
	c	0	8	10	18
	Total	3	97	13	113
	a	7	4	0	11
Teacher	b	2	67	0	69
	c	0	15	19	34
	Total	9	86	19	114
	a	70	2	4	76
Vehicle	b	4	2	3	9
	c	12	0	16	28
	Total	86	4	23	113
	a	16	8	7	31
Green-eyed man	b	3	30	0	33
	c	6	12	28	46
	Total	25	50	35	110
	a	22	4	2	28
Hair–eye	b	1	8	1	10
	c	16	3	55	74
	Total	39	15	58	112
	a	15	6	6	27
Mother–daughter	b	8	4	1	13
	c	9	4	58	71
	Total	32	14	65	111

Note. The entries in the cells are the numbers of subjects who made the forced-choice response indicated by the row and whose estimates were classified as indicated by the column. If the estimate for (b) was larger than that for (a), the answer was classified as (b). If the two estimates were equal, the answer was classified as (c).

ever, performance was slightly better on the estimation section, with 65.2% of the estimates corresponding to the correct alternative as opposed to 57.0% of the forced-choice responses (for a given problem, a subject's data was only included in the analysis if the subject had provided answers for both the forced-choice and estimation versions). It is possible that requiring subjects to given numerical estimates may have resulted in a more careful analysis of the problem.

The estimation data were also analyzed for their "reasonableness." Answers were scored as reasonable if they met the following criteria: for

the fever problem if $p(\text{sick}|\text{fever}) > .75$ and $p(\text{fever}|\text{sick}) < p(\text{sick}|\text{fever})$; for the marijuana problem if $p(\text{marijuana}|\text{heroin}) > ½ > p(\text{heroin}|\text{marijuana})$; for the teacher problem if $p(\text{woman}|\text{tacher}) \geqslant ½ > p(\text{teacher}|\text{woman})$; for the vehicle problem if $p(\text{passenger vehicle}|\text{Datsun}) > ½ > p(\text{Datsun}|\text{passenger vehicle})$; for the green-eyed man problem if $p(\text{man}|\text{green eyes}) = ½ > p(\text{green eyes}|\text{man})$; and for the hair–eye problem if $p(\text{green eyes}|\text{brown hair}) <$ both $p(\text{brown hair}|\text{green eyes})$ and ½. Most of the answers were numerically reasonable if they were ordinally correct, the percentages being 84, 85, 87, 94, 96, and 100% for Problems 1–6, respectively.

The different patterns of responses for the probability and percentage versions of the mother–daughter problem that were observed in the first section of the questionnaire also occurred in the estimation data. For the probability version of the problem, 23 subjects estimated $p(\text{blue-eyed daughter}|\text{blue-eyed mother})$ as greater than $p(\text{blue-eyed mother}|\text{blue-eyed daughter})$, 6 subjects estimated it as less, and 27 subjects gave the same estimates for both conditionals. The corresponding figures for the percentage version of the problem were 8, 9, and 38. These two patterns of responses differ significantly, $\chi^2(2) = 9.83$, $p < .02$, suggesting that the tendency to respond in a way consistent with the existence of a causal bias depends critically on the wording of the problem.

Certain patterns of estimates suggest that some subjects may have confused conditional and joint probabilities. If both estimates were equal and less than 50%, it is possible that both conditional probability expressions were interpreted as the joint probability or that both were interpreted as the smaller conditional probability. The frequency with which this pattern occurred was 3, 5, 1, 3, 21, and 35 for Problems 1–6, respectively, suggesting that one or both of these confusions may have been particularly troublesome for the two attribute problems. Pairs of estimates that were equal but greater than 50% seem less likely to have resulted from a confusion between conditional and joint probabilities, although they could occur if subjects interpreted both expressions as the larger conditional probability. This type of pattern occurred much less frequently: 4, 2, 8, 8, 1, and 3 times for Problems 1–6, respectively.

The estimation data also suggest that certain possible confusions did not play a large role. It is possible, for example, that subjects may have reversed the two conditionals (i.e., interpreted $p(A|B)$ as $p(B|A)$ and vice versa). However, had they done so, we would expect estimates reasonable for $p(A|B)$ to be given for $p(B|A)$. Using the criteria for reasonable estimates given earlier, there were few patterns of estimates of this type: 3, 1, 9, 3, 4, and 4 for Problems 1–6, respectively. Thus with the possible exception of the teacher problem, there was little indication that subjects reversed $p(A|B)$ and $p(B|A)$.

An additional characteristic worth noting about the estimation data was the tendency for subjects to give pairs of estimates that added to 100%. The percentage of subjects doing so for Problems 1–7 were 24, 29, 22, 21, 18, 24, and 44%. Some of these responses may reflect the misconception that pairs of opposite conditional probabilities have to sum to 100%.

Analysis of conditional probability calculations. The conditional probability calculation problem was constructed using the same events (having brown hair and green eyes) as Problem 6, in the hope of getting a better idea where subjects were going wrong on that problem. Three of the 60 subjects left the calculation section blank. On the average, the remaining 57 did better on the calculation problem than they had on the forced choice version: 33 of the 57 subjects (58%) gave the correct answer for both conditional probabilities and an additional 5 subjects clearly had the right idea but made mistakes in either counting or arithmetic (see Table 5). Of the remaining 19 subjects, 14 gave as answers that two *joint* probabilities, whereas only 2 gave responses that suggested that one conditional probability was confused with the other. We do not know whether subjects made the same confusions in providing estimates as they did when calculating. However, as mentioned in the previous section, a common error in the estimation data was for subjects to provide two small and equal numbers that were plausibly estimates of the joint probability.

Subjects' justifications of their answers. Subjects' written justifications of their answers were brief, usually just a sentence and often just a phrase or two. The brevity of the answers made these data less useful than we had hoped. However, the data were adequate to indicate that subjects did use different kinds of justifications for the different problems.

TABLE 5
CALCULATIONS OF CONDITIONAL PROBABILITIES

Calculation of p(green eyes/brown hair)	Calculation of p(brown hair/green eyes)				
	60%	30%	12%	Other	Total
30%	33	0	0	2	35 (61%)
60%	0	2	0	0	2 (4%)
12%	1	0	14	0	15 (26%)
Other	1	1	0	3	5 (9%)
Total	35 (61%)	3 (5%)	14 (25%)	5 (9%)	57

Note. The correct conditional probabilities are 60 and 30% (as indicated by underlining) and the joint probability is 12%. The entries in the cells are the numbers of subjects who gave the answers indicated by the row and column headings.

The most interesting information came from justifications that involved references to causality in the mother–daughter problem. Only 18 subjects justified the judgment that p(blue-eyed daughter|blue-eyed mother) should be greater than the opposite conditional by indicating that a mother could influence the eye color of her daughter but the daughter could not influence the eye color of her mother. However, 15 of these 18 subjects had received the probability version of the problem. If the use of this type of justification was equally likely for both versions of the problem, the probability of a split as extreme as 15/3 would be only .007. This result is consistent with our earlier finding that the forced choice and estimation data for the percentage version of the problem provide little indication of a causal bias in estimating conditional probabilities.

Several additional justifications were mentioned frequently. In the fever problem, the majority of correct answers were justified by asserting that fever implies sickness but sickness does not necessarily imply a fever. Fifty-four subjects made both of these assertions and an additional 34 made one of them. Thirty-four subjects also made similar assertions in justifying their answers to the vehicle problem, even though neither conditional probability in the problem was objectively close to one. However, comments made by some of the subjects indicate that some of these justifications may have followed from a belief that Datsun did not make anything other than automobiles.

For the marijuana problem, 85 subjects included in their justifications phrases such as "marijuana comes before heroin" or "heroin is stronger than marijuana." Although these justifications may have reflected a belief that heroin use implies the use of marijuana but not the reverse, this implication was not stated explicitly in the justifications and only eight subjects gave 100% as their estimate for p(marijuana|heroin).

The protocols did not shed much light on why subjects performed so poorly on Problems 5 and 6. However, for the eye–hair problem, 17 subjects justified their answers that both conditional probabilities were equal by using phrases like "eye color is independent of hair color" or "brown-haired people may or may not have green eyes." These statements may arise from confusions between the notions of "independent events" and "equally probable events."

GENERAL DISCUSSION

Performance varied widely across the different problems, varying from about 80% correct on the implication problems to about 25% correct on Problems 5 and 6. Because subjects did quite well on the implication problems and two of the four real-world knowledge problems in Experiment 2, it appears that statistically naive college students are capable of grasping the concept of conditional probability and its directionality. The very poor performance on Problems 5 and 6 suggest that certain factors can interfere with subjects' basic ability to deal with conditionality. Given the problems we used, we might speculate that such factors include the wording used to express attributes, the difficulty of making the necessary probability estimates on the basis of real-world knowledge, and perhaps confusion between the notions of "independent events" and "equally probable events." However, any definitive statements about these factors would require systematic exploration using many problems varying on a number of dimensions.

Averaged over problems, the percentage of correct responses was almost identical for the probability and percentage versions. However, looked at more closely, there were some clear differences. The probability version led to fewer reversals in Problems 1–6, while the percentage version led to fewer errors of the sort that have previously been considered to be evidence for causal bias. Evidence from both the forced choice and estimation sections of the questionnaire suggests that subjects respond differently to the probability and percentage versions of the mother–daughter problem. The percentage version provided almost no evidence for any effect that might be called "causal bias" although the asymmetry in *a* and *b* responses clearly existed for the probability version. We might speculate that the use of the word *if* in the probability version may have elicited causal reasoning; almost all subjects who provided a clear causal justification of their answer had been given the probability version of the problem. In fact, we cannot rule out the possibility that some of the subjects interpreted the probability version as requiring a judgment about causality.

A distinction that one might wish to draw when students have difficulty with a word problem is whether the underlying cause is some flaw or inadequacy in their underlying conceptual structure (i.e., a misconception or lack of some fundamental concept) or derives from problems in translating from prose to the conceptual structure. Our results are consistent with the hypothesis that difficulties with conditional probabilities are often due to translation errors and suggest that we should be cautious about making strong conclusions about flaws in reasoning or underlying concepts until we have determined how sensitive patterns of errors are to details of the wording.

Our data suggest that translation errors may be reduced if subjects have available some schema in which they can embed real-world events. Subjects' responses and written justifications of their responses seem to suggest the use of a schema such as "if B usually implies A but A can easily occur without B, then $p(A|B) > p(B|A)$" for the fever problem and to a lesser extent the vehicle problem. A similar schema, "if A precedes or is less potent than B, then $p(A|B) > p(B|A)$ seems consistent with the data and written justifications for the marijuana problem.

Finally, our data lead us to suspect that a major translation error may be a confusion between $p(A|B)$ and P(A and B). The most common error in the third part of the questionnaire was to calculate the joint rather the requested conditional probability and the patterns of estimates provided by many subjects in the second section were consistent with what would be expected if they were estimating joint probabilities.

In some sense our results are complementary to the "conjunction effect" discussed by Tversky and Kahneman (1983). Tversky and Kahneman showed that, for example, when statistically naive students made estimates about the results of a health survey conducted on a sample of adult males of all ages and occupations, estimates were higher for the question "What percentage of the men surveyed both are over 55 years old and have had one or more heart attacks?" than for "What percentage of the men surveyed have had one or more heart attacks?" It seems likely that for some subjects this error is related to confusion between the joint and conditional probabilities of the two events. Tverskey and Kahneman found that somewhat higher estimates were given when subjects were asked to estimate the conditional ("Among the men surveyed who are over 55 years old, what percentage has had one or more heart attacks?"). Although this latter result indicates that joints and conditionals were not completely confused, it does not rule out the possibility that this confusion existed for some subjects. One might speculate that instead of separate, differentiated concepts of joint and conditional probability, some subjects may have available a concept that is some amalgam of the two.

REFERENCES

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

Casscells, B. S., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, **299**, 999–1000.

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). New York: Cambridge Univ. Press.

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

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, **101**, 242–254.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgment under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293–315.

RECEIVED: September 6, 1985