Perception of correlation reexamined

RUTH BEYTH-MAROM
Decision Research, Perceptronics, Eugene, Oregon 97401

Almost all studies of adult notions of correlation between dichotomous variables show that people do not incorporate two conditional probabilities as they should according to normative definitions. However, these studies disagree considerably about what correlational notions people do have. This paper identifies three factors that contribute to the variability in research results. The first two factors were mentioned in the literature, and the evidence concerning them is summarized: (1) the way data are presented and (2) the instructions subjects receive. A third factor is suggested and studied; the type of variables between which correlation is judged may affect subjects' notion of correlation. Specifically, asymmetric, present/absent variables (e.g., symptom: present, absent) may strengthen the incorrect notion of correlation as the tendency of two events to coexist (e.g., presence of symptom and presence of disease) disregarding the complementary events. In three experiments, subjects were asked to choose among five interpretations of the sentence "A strong [or no] relationship exists between [two variables]." The above prediction was confirmed.

Adults' perception of the correlation between dichotomous variables had been examined in a number of studies using a common paradigm: Subjects must identify the strength and/or direction of the relationship between two variables on the basis of many pairs of data. These pairs can be summarized as in Table 1, in which A_1 and A_2 are the two values of Variable A, B_1 and B_2 are the two values of Variable B, and a, b, c, and d represent cell frequencies.

The statistical correlation between A and B is a function of the difference between two conditional probabilities: $P(B_1/A_1) = a/(a + c)$ and $P(B_1/A_2) = b/(b + d)$. The form of the function depends upon the particular correlation coefficient one chooses to calculate (Hunter, 1973; Särndal, 1974).

Most research evidence (reviewed by Crocker, 1981; Shaklee & Tucker, 1980) shows that people do not have the "right" perception of correlation: Perceived correlation is not a function of these conditional probabilities. However, investigations using the above paradigm disagree considerably about what it is a function of. Smedslund (1963) concluded that "subjects' strategies and inferences typically reveal a particularistic non-statistical approach or an exclusive dependence on the frequencies of the [a] instances" (p. 172). Inhelder and Piaget (1958) identified this as the strategy used by younger adolescents (12-13 years). Ward and Jenkins (1965) found that subjects rely on Cell d as well as Cell a, making their focus the "confirming" cases. Shaklee and Tucker (1980), however, demonstrated that many subjects compare the number of confirming cases

Table 1
Number of Cases (Out of N) in Which the Possible Values of Variables A and B Cooccur

Values of Variable B	Values of Variable A		
	A_1	A_2	Sum
B_1	a	b	a+b
B_2	c	d	c+d
Sum	a+c	b+d	

(a and d) to the number of disconfirming cases (b and c), thus attending to all four cells of the 2 by 2 table. This notion of correlation develops, according to Inhelder and Piaget, at the age of 14-15 years and is characteristic of formal operational thought.

In a study comparing judgment of contingency of depressed and nondepressed students, Alloy and Abramson (1979) demonstrated that "at least under some circumstances, when a contingency between responses and outcomes exists, subjective representations of contingencies mirror objective contingencies across a wide range of the response-outcome contingency space" (p. 455). This was true for depressed as well as nondepressed subjects. However, the nature of the above experiments does not enable one to determine the rule subjects were using.[1]

I would like to propose three explanations for the variability in the results. Two of these explanations were mentioned and partly studied in the literature, whereas the third has been ignored. Some empirical data concerning the third explanation will be presented.

Sources of Variability

Data presentation. In Smedslund's (1963) first experiment, in Jenkins and Ward's (1965) sole experi-

ment, and in one of Ward and Jenkins' (1965) conditions, the data were presented serially, trial by trial. After observing the data once, subjects judged the strength of relationship between the two variables on either a 7- or a 10-point scale ranging from "perfect relationship" to "no relationship." According to Hamilton (1981), this task involves four steps: (1) encoding the relevant information, (2) remembering the relevant information, (3) retrieving that information, and (4) integrating the information into a judgment. The last step is actually composed of two different steps: (4a) organization (into a 2 by 2 table or other arrangement) and (4b) assessment of correlation. The tendency to overrespond to salient components of the stimulus field (Taylor & Fiske, 1978) and the failure to recognize the relevance of nonoccurrences (Hamilton, 1981) may bias the encoding, retention, and retrieval processes, thereby affecting indirectly the last integration stage of information organization and assessment of correlation.

Alert to this possibility, Smedslund (1963), in a second experiment, gave subjects the list of pairs, thus aiding their memory. This manipulation did not improve their performance. In one of their conditions, Ward and Jenkins (1965) eliminated the first three stages plus half of the fourth one by presenting the data in a 2 by 2 table. This summary presentation improved subjects' performance substantially over a trial-by-trial presentation. However, surprisingly, being exposed to a 2 by 2 table after seeing a trial-by-trial presentation did not have the same upgrading effect. All of Shaklee and Tucker's (1980) subjects had all relevant available material on typed cards when judging correlations. One group was explicitly encouraged to sort the data into a 2 by 2 table, whereas the other one was not. No performance differences were detected between these two groups, probably because 50% of the subjects who received no sorting instructions nevertheless spontaneously sorted the cards.

Table 2 summarizes the ways in which different data presentations may change both task characteristics and the demands on subjects. Trial-by-trial presentation is the most difficult, because it requires subjects to memorize, organize, and judge. Listing stimuli requires organization and judgment, whereas tabular presentation requires judgment only.

Type of instruction. None of the reported experiments asked subjects simply to "judge the correlation (or relationship)" between the two variables. Rather, each explained the task in some detail. Evidently, the experimenters felt either that subjects are not acquainted with the concepts of "correlation" and "relationship" or that technical and lay usage of these concepts are very different. However, these explanations and instructions (and, therefore, the task) differed from experiment to experiment (see Table 2, Column 3).

In Smedslund's (1963) experiments, subjects saw 100 cards, on each of which were written two letters. The upper part of the card listed a "symptom" and the lower part, a "diagnosis." The symptom was one of the letters A, B, C, D, or E; the diagnosis was one of the letters F, G, H, I, or J. The instructions read: "We are interested in learning how well beginner students of nursing are able to form an opinion about the practical usefulness of symptoms in diagnosis . . . . You are to concentrate entirely on symptom A and diagnosis F. Your task is to look through the pack of cards once and form an impression of the extent to which A is a useful symptom in the diagnosis F. In other words, do you think A is a symptom one should pay attention to in trying to determine whether or not the patient is likely to be diagnosed as F?" (Smedslund, 1963, p. 164).

These instructions seem to have focused subjects' attention on Cell a, the (A,F) pair, and diverted attention from the other cells, $(\bar{A},F)$, $(A,\bar{F})$, and $(\bar{A},\bar{F})$. If this is the case, then Smedslund's (1963) subjects did exactly what they were told to do, which was not to judge the relationship between two dichotomous variables, but to judge Cell a's frequency relative to N, the number of stimuli presented.

Ward and Jenkins (1965) gave subjects information

Table 2
A Comparison Between Studies on the Perception of Correlation

	Form of Data Presentation	Instructions Pointing at	Type of Variables	Results (Strategy)
			Smedslund (1963)	
Experiment 1	trial by trial	a	asymmetric	a
Experiment 2	list of data available	a	asymmetric	a
			Jenkins and Ward (1965)	
	trial by trial	a	1 of 2 variables symmetric, alternative not	a
			Ward and Jenkins (1965)	
Condition D	trial by trial	a+d	asymmetric	a+d
Condition DT	trial by trial; table	a+d	asymmetric	a+d
Condition T	table	a+d	asymmetric	correct one
			Shaklee and Tucker (1980)	
Experiment 1	list of data available	a,b	symmetric and asymmetric	(a+d) − (b+c)
Experiment 2	list of data available	a,b,c,d	symmetric and asymmetric	(a+d) − (b+c) and correct one

concerning whether or not cloud seeding had taken place and whether or not rain had fallen in different states of the U.S. The instructions were: "At the end of the experiment for each state, you are to judge how much control seeding the clouds had over the occurrence of rainfall in the state . . . . Complete control means that whenever you seed, it rains, and whenever you don't seed, it does not rain" (Ward & Jenkins, 1965, p. 235). Whereas Smedslund's (1963) instructions pointed directly at Cell a, Ward and Jenkins' pointed at the confirming cases of Cells a and d. As noted, subjects based their judgments on exactly these cells.

In Shaklee and Tucker's (1980) experiments, correlational problems were structured in such a way that patterns of correct and incorrect judgments would indicate the judgmental strategy being used by each subject (the strategies being the names of the columns in Table 3). Their first experiment presented individual data instances on index cards. The instructions read as follows: "Use these cards to decide whether this plant is one that stays healthiest when given a large glass of water each week, or when given a small glass of water each week, or doesn't it make any difference to the plant?" (Shaklee & Tucker, 1980, p. 462). In their discussion, Shaklee and Tucker observed: "A subject who is asked if Outcome A_1 is associated with B_1 or B_2 may simply compare the frequencies in the cells with those particular event combinations ($A_1 B_1 - A_1 B_2$). With problem structures used in this experiment, the comparison would be between Cells a and b in a traditionally labeled contingency table (a vs. b strategy)" (1980, p. 464). Hence, these instructions encourage the subject to compare Cell a to Cell b. The top row of Table 3 shows the distribution of subjects according to the strategies they used.

In their Experiment 2, Shaklee and Tucker (1980) chose items that enabled them to differentiate among all of the various strategies, and they presented the data in a 2 by 2 table. One group received the Experiment 1 instructions, whereas the other group received a response form that asked them about the relative likelihood of an event (A_1), given each of the alternative states of the second event (B_1 and B_2). An example question is shown in the appendix.

These latter instructions directed subjects to the correct interpretation of correlation (the difference between conditional probabilities). These differential instructions should have altered the distribution of the subjects' preferred strategies, producing a high percentage of the "a vs. b" strategy in the "a vs. b" instructions and a high percentage of the correct strategy with the "conditional" instructions. These two distributions are shown in the second and third rows of Table 3. As expected, there were more "a vs. b" strategies under the "a vs. b" instructions and more correct strategies under the conditional instructions. Improved performance of the "a vs. b" group in Experiment 2 (compared with Experiment 1 subjects) may be explained by the difference in data presentation. Fairly good performance was also observed in Alloy and Abramson's (1979) study, in which instructions explaining the concept of control were extremely detailed (over 300 words).

In short, subjects judging relationship appear to do what they are told to do. As a result, different instructions lead to different behavior.

Symmetric vs. asymmetric variables. A third variable that may affect the perception of correlation is the type of variables that are chosen for the study. One important distinction is between (1) asymmetirc variables such as "symptom" (symptom present, symptom absent) and (2) symmetric variables such as "gender" (male, female).

The symmetric-asymmetric differentiation relates to the differential vs. similar status of the two variables' values. In the asymmetric case, the "absent" value has a lower status than the "present" value, whereas in the symmetric case, both values have a similar status. Evidence for this differentiation is semantic and empirical. In the asymmetric case, the name of the variable is like the name of one of its two values, (e.g., the variable is "pneumonia"; the values are "pneumonia" and "no pneumonia"). This is very similar to the unmarked vs. marked differentiation of adjectives (Clark, 1969; Greenberg, 1966), in which the unmarked member of a pair of adjectives (high in high-low) serves as the measure of the full scale (height). With symmetric variables, the name of the variable ("gender") differs from the name of its two values ("male," "female"). Furthermore, in the asymmetric case, the two values may be described as "occurrence," "nonoccurrence" or "positive," "negative." A "nonoccurrence" or a "negative" event has

Table 3
Subjects' Distribution Across Strategies in the Study by Shaklee and Tucker (1980)

Instructions Pointing at		Chosen Strategy				
		a	a vs. b	(a+d) − (b+c)	Correct	Unclassified
Experiment 1	a vs. b	17.0		41.0	17.0	23.0
Experiment 2	a vs. b	2.1	22.4	25.1	23.4	17.0
	Conditional Probabilities	.0	13.3	34.5	42.2	10.0

Note−Unfortunately, for the specific stimuli used in Experiment 1, the a vs. b strategy resulted in the same response pattern as did the (a+d) − (b+c) strategy (i.e., comparing confirming cases with disconfirming cases). The data in Experiment 2 did not appear in Shaklee and Tucker (1980); they are reported here by permission of Shaklee and Tucker. Values are expressed as percentages.

less impact on people's attention than a positive event (Nisbett & Ross, 1980). Thus the two values are weighted differently.

A correlation between two asymmetric variables, such as symptom and disease, may be easily interpreted by subjects as the tendency of the two "present" values (presence of symptom and presence of disease) to coexist, thus interpreting relationship between variables as a relationship between values. This interpretation of correlation seems to fit the common language interpretation that the *American Heritage Dictionary of the English Language* (1975) defines as "a logical or natural association between two things, relevance of one to another; connection." An intuitive interpretation of correlation as a relationship between variables is more likely with symmetric variables such as sex (male, female) and height (tall, short). In the symmetric case, people probably ask themselves whether most females are short and most males are tall. One would, therefore, expect more serious misinterpretations of correlation in the asymmetric case than in the symmetric one, which will express itself in attention to fewer cells of the 2 by 2 table. This prediction is reinforced as well by the differential weight people give to the two values in the asymmetrical case.

Variable type has not been studied previously. Smedslund's (1963) subjects' poor performance (reliance on Cell a only) was demonstrated with asymmetric variables. However, this may be due to instructions that pointed at Cell a. Ward and Jenkins (1965) used asymmetric variables, too, but their instructions directed subjects strongly to Cells a and d. Although Shaklee and Tucker (1980) used both asymmetric and symmetric pairs of variables, these pairs were not analyzed separately.

EXPERIMENT 1

An experiment was conducted (1) to find out how subjects intuitively interpret correlation when given minimal instructions and (2) to test the hypothesis that intuitive perception of correlation is a function of the type of variable. Specifically, it was predicted that with asymmetric variables there will be a stronger tendency to interpret correlation as a tendency of the two "present" values to coexist, thus attending to fewer cells of the 2 by 2 table (compared with the perception of correlation between two symmetric variables).

This predicted tendency may be a function not only of the type of variable, but also of the extent to which the two values of each variable are specified. Explicit specification of the "present" and "absent" values of the asymmetric variables may weaken this tendency and cause subjects to attend to the "absent" values as well.

Method

In previous experiments, subjects were asked to assess the correlation for a given set of data. From the assessed correlations, the researcher inferred their intuitive perception of correlation (i.e., which cells in a 2 by 2 table were perceived as being relevant). The following experiment differed in two ways: (1) The stimuli were not the raw material (pairs of data), but a given correlation ("a strong relationship exists between . . . "), and (2) subjects indicated directly their intuitive perception of correlation. Specifically, subjects chose among five possible interpretations of the phrase "there is a strong relationship between [two variables]." Previous results guided the composition of the five possible interpretations that were offered. These can be defined by the different cells of a 2 by 2 table: a, b − a, a + d, (a + d) − (b + c), and the "correct" one.

Four variables were chosen: two symmetric (skin color: dark, light; temperature: high, low) and two asymmetric ones (symptom: present, absent; disease: present, absent).

Questionnaires. To test the effect of variable type (symmetric vs. asymmetric) and value specification (values specified vs. values not specified) four different one-page questionnaires were constructed. Each had an introductory sentence, an instruction sentence, and five possible interpretations. The introductory sentence read as follows: for the symmetric-specified questionnaire, "A paper published in a major biological journal reported that for one species of widely distributed animals *a strong relationship was found between the animal's skin color (light/dark) and the mean temperature (high/low) in its territory,*" and for the asymmetric-specified questionnaire, "A paper published in a major medical journal reported that for one species of animals *a strong relationship was found between a specific symptom (exists/does not exist) and a specific disease (exists/does not exist).*" In the two nonspecified questionnaires, the introductory sentences were similar except for the omissions of the parentheses' content.

The instruction sentence was identical in all four questionnaires: "Below are five different interpretations of the underlined clause. Read all of the interpretations carefully. Be sure you understand each of them. Please choose *one* that fits your interpretation best and circle its number."

The five interpretations were similarly phrased in all four questionnaires, except for essential content differences. The following are from the symptom-disease questionnaires. "(1) Among all animals examined, many had the symptom *and* the disease. (2) Among all animals examined with the disease, there was a higher percentage with the symptom than without it. (3) Among all animals examined, there were many animals either with the disease and the symptom or without both. (4) Among all animals examined, there was a higher percentage of either animals with the symptom and the disease or animals without both *than* either animals with the disease, but without the symptom or animals without the disease but with the symptom. (5) The percentage of animals with the symptom among animals with the disease was higher than the percentage of animals with the symptom among animals without the disease."

The five interpretations are, respectively, a, b − a, a + d, (a + d) − (b + c), and the "correct" one. The order of the five interpretations was varied within each group to balance order effects.

The last sentence in each questionnaire was "*After* you have chosen one of the above, if you have a different interpretation, write it below." Only a few responses were offered to the last sentence, all of these in the color-temperature groups. In these, subjects correctly indicated that the introductory sentence does not point at the direction of the relationship, only at its strength.

Subjects. Subjects were 273 paid volunteers who responded to an ad in the University of Oregon student newspaper. The questionnaires were randomly distributed among all subjects. The questionnaire was self-paced and embedded in a 120-min experimental session involving a variety of unrelated judgment tasks.

Results

The frequency distribution of subjects along the five different interpretations for the four groups is given in the upper part of Table 4.

Specification of values. There was no difference between the "specified" and the "nonspecified" conditions. Chi-square tests for the difference were not significant in either the "symptom-disease" or the "color-temperature" cases [$\chi^2(4) = 4.91$ and 3.29, respectively]. Therefore, the two conditions were combined (the combination of Columns 1 and 3 resulted in Column 5, and the combination of Columns 7 and 9 resulted in Column 11).

Type of variable. A chi-square test between Columns 5 and 11 revealed a significant "type-of-variable" effect [$\chi^2(4) = 32.26$, $p < .001$]. To clarify the difference between the symmetrical and asymmetrical cases, the first two "narrow" interpretations of correlation were combined (Rows 1 and 2), as well as the next two, "broader" ones (Rows 3 and 4). The left portion of Table 5 is the result of these aggregations [$\chi^2(2) = 32.07$, $p < .001$]. In both conditions, only about one

subject in eight chose the correct interpretation. However, their errors were quite different: About one-third of the subjects in the asymmetirc condition interpreted relationship in a very narrow way, whereas less than 6% of subjects in the symmetric condition did so. In both conditions, the majority of subjects interpreted correlations as either "the number of confirming cases" or "the difference between number of confirming cases and the number of disconfirming cases."

Discussion

Past evidence that subjects do have a notion of correlation was restricted to a very specific situation in which they were explicitly instructed to judge the relation between $P[a/(a + c)]$ and $P[b/(b + d)]$ (Shaklee & Tucker, 1980). The present results indicate that when subjects are simply asked how they interpret "strong relationship between variables," they do not choose the correct statistical interpretation from among those presented to them. Most people's intuitive notion of correlation is different from the statistical normative one.

Table 4
Frequency Distributions Over the Five Interpretations for Four Groups

	Symmetric Variables						Asymmetric Variables					
	Specified		Nonspecified		Combined		Specified		Nonspecified		Combined	
Strategies	Raw 1	% 2	Raw 3	% 4	Raw 5	% 6	Raw 7	% 8	Raw 9	% 10	Raw 11	% 12
				Experiment 1: Strong Relationship								
a	1	1.5	2	2.8	3	2.2	11	15.1	8	12.9	19	14.1
(b−a)	1	1.5	4	5.6	5	3.6	10	13.7	15	24.2	25	18.5
(a+d)	21	31.3	20	28.2	41	29.7	17	23.3	13	21.0	30	22.2
(a+d) − (b+c)	33	49.3	38	53.5	71	51.5	29	39.7	17	27.7	46	34.1
Correct	11	16.4	7	9.9	18	13.0	6	8.2	9	14.5	15	11.1
Total	67	100.0	71	100.0	138	100.0	73	100.0	62	100.0	135	100.0
				Experiment 2: No Relationship								
a	2	3.0	2	3.2	4	3.1	15	25.0	11	18.9	26	22.0
(b−a)	3	4.5	10	16.1	13	10.1	14	23.3	8	13.8	22	18.7
(a+d)	6	9.0	4	6.5	10	7.8	3	5.0	0	.0	3	2.5
(a+d) − (b+c)	48	71.6	35	56.5	83	64.3	8	13.3	19	32.8	27	22.9
Correct	8	11.9	11	17.7	19	14.7	20	33.3	20	34.5	40	33.9
Total	67	100.0	62	100.0	129	100.0	60	99.9	58	100.0	118	100.0

Table 5
Frequency Distributions Over Three Combined Interpretations for the Two Combined Groups

	Experiment 1: Strong Relationship				Experiment 2: No Relationship			
	Symmetric		Asymmetric		Symmetric		Asymmetric	
Strategies	Raw 1	% 2	Raw 3	% 4	Raw 5	% 6	Raw 7	% 8
a (b−a)	8	5.8	44	32.6	17	13.2	48	40.7
(a+d) (a+d) − (b+c)	112	81.2	76	56.3	93	72.1	30	25.4
Correct	18	13.0	15	11.1	19	14.7	40	33.4
Total	138	100.0	135	100.0	129	100.0	118	100.0

What notion do they have? Most subjects believe that correlation is a function of all available information (all four cells of a 2 by 2 table), but they aggregate it incorrectly. This is a more developed notion of correlation than the one relying on only part of the information [a or (b − a)]. When $A_1 = A_2$ and $B_1 = B_2$, one will arrive at the same numerical correlation using the $[(a + d) − (b + c)]/n$ strategy or the correct strategy based on the difference between conditional probabilities. Furthermore, when only one of the marginal distributions is uniform, the $[(a + d) − (b + c)]/n$ strategy will yield the same result as the correct strategy when the conditional events are those that are evenly distributed.

The less developed notion of correlation, believing correlation is a function of only one or two cells of a 2 by 2 table, is much more popular (33% vs. 6%) when the variables involved are asymmetric (present/absent variables) as compared with symmetric ones. This last result supports the hypothesis that with asymmetric variables, there is a tendency to confuse variables with one of their values, interpreting relationship as the tendency of two events (values) to coexist without regard for the complementary events. This misinterpretation did not diminish even when the two possible values (exists/does not exist) were clearly stated.

The present results are, however, restricted to subjects' interpretation of "strong relationship." The obvious next step is to see how subjects interpret "no relationship" and whether their interpretation similarly depends on the type of variable specified.

EXPERIMENT 2

Method

The four Experiment 1 questionnaires were minimally changed; in the introduction sentences, "a strong relationship" was changed to "no relationship." The instruction sentences were unchanged. The five interpretations were similarly phrased in all four questionnaires except for essential content differences. The following are from the symptom-disease questionnaire: "(1) Among all animals examined, only few had the symptom *and* the disease. (2) Among all animals examined with the disease, there was the same number with the symptom as without it. (3) Among all animals examined, there were few animals either with the disease and the symptom or without both. (4) Among all animals examined, there was the same percentage of animals either with the symptom and the disease or without both *as* animals either with the disease but without the symptom or without the disease but with the symptom. (5) The percentage of animals with the symptom among animals with the disease was the same as the percentage of animals with the symptom among animals without the disease."

The five interpretations are, respectively, "few a" (instead of "many a" in Experiment 1), b = a (instead of b − a), "few a + d" (instead of many a + d), (a + d) = (b + c) [instead of high (a + d) − (b + c)], and the correct one (equality of two conditional probabilities instead of a difference between them). The five interpretations were presented in five different orders to control for order effects.

Subjects. Subjects were 247 paid volunteers recruited and tested in the same manner as those in Experiment 1. The questionnaires were randomly distributed among all subjects.

Results

The frequency distribution of subjects' responses to the five different interpretations for each of the four groups is given in the lower part of Table 4.

Specification of values. No significant difference was observed between the "specified" and "nonspecified" conditions in the color-temperature conditions $[\chi^2(4) = 6.99]$. Although the difference between the two conditions in the symptom-disease case was significant $[\chi^2(4) = 9.64, p < .05]$, its direction was opposite to the one predicted prior to Experiment 1, namely, that more serious misinterpretations would occur in the "nonspecified" than in the "specified" case. Furthermore, the difference was only slightly beyond the point of significance ($\chi^2 = 9.49$ at the .05 level). Hence the results of the "specified" and "nonspecified" conditions were combined.

Type of variable. A chi-square test between Columns 5 and 11 revealed a significant "type-of-variable" effect $[\chi^2(4) = 57.81, p < .001]$. As in Experiment 1, the first two interpretations were combined as well as the next two (see left portion of Table 5). A "type-of-variable" effect was manifested again $[\chi^2(2) = 54.15, p < .001]$. As in Experiment 1, the interpretation of "no relationship" in a very narrow way was more popular in the asymmetric conditions (40%) than in the symmetric ones (13%). Thus, subjects in the latter conditions performed better than those in the former ones. In contrast to Experiment 1, subjects in the asymmetric conditions performed better than those in the symmetric ones in the sense that more subjects in the former groups chose the correct interpretation (33.9% vs. 14.7%; z = 3.55, p < .001).

Comparing the results of the two experiments (see Table 5), one finds no significant difference between the color-temperature conditions in the two experiments $[\chi^2(2) = 4.69]$ but a significant one between the symptom-disease conditions $[\chi^2(2) = 30.36, p < .001]$. Under the symptom-disease conditions, "no relationship" was interpreted correctly more often than was "strong relationship."

Discussion

The results of Experiment 1 regarding the perception of strong relationship now can be generalized to the perception of no relationship. Subjects misinterpret the concept of relationship more often when the considered variables are asymmetric than when they are symmetric.

The type of variable is sometimes a question of formulation more than a description of the underlying phenomenon; hence, one can often choose how it is labeled. For example, in checking patients with tumors, one can either check for the presence or absence of malignant tumors, or for the type of tumor: malignant or benign. This freedom of formulation is more limited in other cases, like the variables chosen by Ward and Jenkins (1965): seeding/not seeding and rain/no rain. It is difficult (if not impossible) to reformulate an

asymmetric variable that describes the presence or absence of an action.

The third experiment was designed to test whether changes in the formulation (symmetric vs. asymmetric) of very similar variables can affect the interpretation of the relationship concept. It also served the purpose of replicating the first two experiments with a different pair of variables.

EXPERIMENT 3

Method

Two pairs of variables were chosen: In one pair, the two variables were symmetric, and in the other, they were asymmetric. However, the two pairs were very similar. The first variable was "a pigment" (present/absent) in the asymmetric case and "pigmentation" (dark/light) in the symmetric one. The second variable was "social behavior" (present/absent) in the asymmetric condition and "peer behavior" (cooperative/competitive) in the symmetric one.

Four different questionnaires were prepared, differing in the type of variable (symmetric vs. asymmetric) and the strength of relationship (strong relationship vs. no relationship). All four questionnaires specified the two values of the variable.

The first sentence of the two symmetric questionnaires read as follows: "A paper published in a major journal reported that for one species of mice *a strong relationship (no relationship) was found between the pigmentation (dark/light) of the mice and their peer behavior (cooperative/competitive)*."

The first sentence of the two asymmetric questionnaires read as follows: "A paper published in a major journal reported that for one species of mice *a strong relationship (no relationship) was found between a pigment (present/absent) and their social behavior (present/absent)*."

The instructions and the structure of the five interpretations were similar to those presented in Experiments 1 and 2. Order effects were controlled by manipulating the order of the five interpretations.

Subjects. Subjects were 334 paid volunteers recruited and tested in the same way as those in the previous experiments. The questionnaires were randomly distributed among all subjects.

Results

Table 6 presents the frequency distribution of subjects along the five different interpretations for the four groups.

A chi-square test comparing the performance under the symmetric and asymmetric variables for the strong relationship conditions (Columns 1 and 3) did not reveal a significant effect. However, when the first two interpretations were combined, as well as the third and the fourth (see Table 7, Columns 1 and 3), the effect was significant $[\chi^2(2) = 6.7, p < .05]$. More subjects under the asymmetric condition chose the more primitive interpretations.

Similar analyses were performed for the "no-relationship" conditions. A chi-square test on Columns 5 and 7 of Table 6 was significant $[\chi^2(4) = 17.6, p < .01]$, as well as a chi-square test on Columns 5 and 7 of Table 7 $[\chi^2(2) = 16.0, p < .001]$. The type-of-variable effect in the no-relationship conditions was not due solely to the higher percentage of correct choices in the asymmetric group (20.4% vs. 8.2%); the same effect was also present when the chi-square test was performed on

Table 6
Frequency Distributions Over the Five Interpretations–Experiment 3

| | Strong Relationship | | | | No Relationship | | | |
| | Symmetric | | Asymmetric | | Symmetric | | Asymmetric | |
Strategies	Raw 1	% 2	Raw 3	% 4	Raw 5	% 6	Raw 7	% 8
a	3	3.8	7	8.6	0	.0	6	6.8
(b−a)	7	8.9	17	20.7	7	8.2	15	17.1
(a+d)	21	26.6	19	23.2	2	2.4	2	2.3
(a+d) − (b+c)	34	43.0	27	32.9	69	81.2	47	53.4
Correct	14	17.7	12	14.6	7	8.2	18	20.4
Total	79	100.0	82	100.0	85	100.0	88	100.0

Table 7
Frequency Distributions Over Three Combined Interpretations–Experiment 3

| | Strong Relationship | | | | No Relationship | | | |
| | Symmetric | | Asymmetric | | Symmetric | | Asymmetric | |
Strategies	Raw 1	% 2	Raw 3	% 4	Raw 5	% 6	Raw 7	% 8
a (b−a)	10	12.7	24	29.3	7	8.2	21	23.9
(a+d) (a+d) − (b+c)	55	69.6	46	56.1	71	83.6	49	55.7
Correct	14	17.7	12	14.6	7	8.2	18	20.4
Total	79	100.0	82	100.0	85	100.0	88	100.0

the four "wrong" interpretations only [for Table 6, $\chi^2(3) = 12.8$, p < .01; for Table 7, $\chi^2(1) = 10.8$, p < .01].

GENERAL DISCUSSION

Taken together, the results of Experiments 1-3 demonstrate a type-of-variable effect on the interpretation of relationship. The perception of relationship as a function of Cell a or Cells a and b is more frequent when the related or unrelated variables are asymmetric (i.e., present/absent variables). It seems that with such variables, more subjects tend to perceive a relationship between two variables as the tendency of the "present" values of both variables to coexist. This simplification may be at least partly due to the difficulty people have in processing negation (Clark, 1974; Wason, 1959).

Different labeling of the same variables may thus affect the way subjects interpret the concept of relationship and, thereby, their judgments of its strength. As such, the labeling of variables can be seen as another demonstration of a "framing effect"; the way in which a problem is presented affects the way the task is performed (Lichtenstein & Slovic, 1971; Tversky & Kahneman, 1981).

The present experimental design was adopted to avoid any possible effect due to type of instructions and method of data presentation. Subjects' perception of relationship was not inferred from their relationship judgments of a set of 2 by 2 tables, as was done in previous research, but was directly observed in their interpretations of the concepts "strong relationship" and "no relationship."

When interpreting "no relationship," between 60% (for the asymmetric variables in Experiment 2) and 90% (for the symmetric variables in Experiment 3) of the subjects chose an interpretation that relates to all cells of a 2 by 2 table [the (a + c) − (b + d) strategy or the correct one]. This is surprisingly good performance. When interpreting "strong relationship" between 45% (for the asymmetric variables in Experiment 1) and 65% (for the symmetric variables in Experiment 1) of the subjects chose such interpretations. Although a smaller percentage, it is still indicative that a substantial percentage of subjects have a perception of relationship that is identical or similar to the statistical concept of correlation.

Two important lessons may be learned from the comparison between the traditional research design and the present one: (1) When we want to study how people interpret a concept, it may help to ask them about it directly. (2) We can also choose different experimental manipulations that will enable us to derive their understanding from the way they respond to different data. In doing so, we have to be very alert to possible task characteristics that may affect subjects' performance.

REFERENCES

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

American Heritage Dictionary of the English Language. New York: American Heritage, 1975.

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, 76, 387-404.

Clark, H. H. Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12): *Linguistic and adjacent arts and sciences.* The Hague: Mouton, 1974.

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

Greenberg, J. H. *Language universals.* The Hague: Mouton, 1966.

Hamilton, D. L. Illusory correlation as a basis for stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior.* Hillsdale, N.J: Erlbaum, 1981.

Hunter, A. A. On the validity of measures of association: The nominal-nominal, two by two case. *American Journal of Sociology*, 1973, 79, 99-109.

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

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

Lichtenstein, S., & Slovic, P. Reversals of preference between bids and choices in gambling situations. *Journal of Experimental Psychology*, 1971, 89, 46-55.

Nisbett, R., & Ross, L. *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, N.J: Prentice-Hall, 1980.

Särndal, C. E. A comparative study of association measures. *Psychometrika*, 1974, 39, 165-187.

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

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

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1978.

Tversky, A., & Kahneman, D. The framing of decisions and the psychology of choice. *Science*, 1981, 219, 453-458.

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

Wason, P. C. The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, 1959, 11, 92-107.

NOTE

1. Subjects in Alloy and Abramson's experiments participated in a 40-trial experiment in which they could either press a button or withhold from pressing. For each response, an outcome was either present (an onset of a green light) or absent. The subjects' task was to learn the contingency between their response (press, not press) and the outcome (green light present, absent). As the subjects were strongly encouraged to try both responses, the probability of a uniform distribution of the responses [i.e., (a + b) = (c + d)] cannot be ignored. Under such conditions, some possible heuristics, wrong as well as right ones, may result in an accurate contingency evaluation. Specifically, the difference between the two conditional probabilities $[a/(a + b) − c/(c + d)]$ is equivalent to the difference between confirming and disconfirming cases $[(a + d)/n − (b + c)/n]$.

Appendix
An Example Question

The picture indicates that when it was snowing blockheads were

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

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

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