

Intuitive covariation estimation

ROBERT A. MALMI

GTE Laboratories, Waltham, Massachusetts

Six experiments concerned people's ability to estimate the degree and sign of covariation represented in a bivariate distribution of stimuli with which they had just been presented as a series of pairs of stimuli. The stimuli were pairs of numbers, pairs of lines of variable lengths, or word-line pairs. In the latter case, subjects were asked to think of the words in terms of either pleasantness or familiarity; hence, the covariation relationship was between the normative pleasantness or familiarity value of the word and a line of variable length. In the sixth experiment, subjects were presented with two word-line pairs and were asked to assess the covariation of both simultaneously. In most cases, the estimates reflected the sign and degree of covariation of the stimuli quite well. The estimates did not reflect accurately the stimulus covariation when the stimuli were numerical and the sign of the correlation of the stimuli was negative. A distinction is made between intuitive and strategic processes in the assessment of covariation.

The ability to predict and estimate characteristics of people, events, and things would seem to be essential in everyday interactions. Knowing that certain behaviors of other people tend to be associated with other behaviors allows one to prepare for the future and behave appropriately. One way to view predictive ability and the knowledge upon which it is based is to consider that it results in part from assessments of the covariation of stimuli derived from experience. That is, one learns from experience that certain kinds of social signals from another person are associated with aggression, affection, or other tendencies. One could view these signals as being correlated with the tendencies. As with most correlations, the relationship is not perfect and has some uncertainty associated with it.

Previous research indicates that people are generally quite poor at judgments about the actual degree of covariation represented in stimuli they inspect (Nisbett & Ross, 1980). Studies have been reported in which subjects were shown pairs of correlated stimuli, such as letters, representing symptoms and diagnoses (Smeldelund, 1963), mood and the weather (Shaklee & Mims, 1982), and so on. Sometimes the stimuli have been presented in tabular form (see Beyth-Marom, 1982, for a summary). Another popular method used to study covariation assessment has been to show subjects a number, ask for a prediction about the value of its pair, and then show the correlated pair as feedback (Birnbbaum, 1976; Brehmer,

1973a, 1973b, 1974; Slovic, 1974). These studies have concluded that people cannot accurately assess the covariation of the stimuli they have inspected.

Nisbett and Ross (1980) contended that people's inability to accrue knowledge about covariation contributes greatly to faulty decision making. They offered little hope for any remedy for this cognitive weakness, inasmuch as accurate covariation estimation "requires mental operations that are simply beyond our intuitive, informal capacities" (p. 286). As will be seen, the major conclusion of the present study is that this is not the case under circumstances in which people rely on intuitive processes to assess covariation and in which inappropriate strategies are discouraged.

The experimental paradigm employed here resembles that used by others (e.g., Birnbbaum, 1976; Brehmer, 1973a, 1973b, 1974; Slovic, 1974) in some respects, but differs in that the procedures were such that inappropriate strategies for assessing the covariation of the stimuli were discouraged. In the usual paradigm, a number is presented, the subject makes a prediction about the value of its pair, and then the other member of the pair is shown. Slovic (1974) asked his subjects how they generated their responses. The reported strategies included adding a constant to the first stimulus, dividing it by something, multiplying, and so on. These strategies do not seem to conform to any of the three senses of the concept *intuitive* described by Kahneman and Tversky (1982, p. 494): informal and unstructured reasoning, a fact or rule compatible with a lay model, and a conduct or procedure that is part of normal reasoning. Thus, it seems reasonable to speculate that intuitive processes may not have been measured in the studies of covariation estimation listed above.

A modification of this experimental paradigm is introduced in Experiment 1. As in the many studies using the stimulus-response-feedback procedure discussed above, the stimuli shown to subjects in Experiment 1 were

This research was supported by Grant No. MH 37166 from the National Institute of Mental Health. The data were collected when the author was on the faculty at the State University of New York at Binghamton. Dan Burns and David Samson helped with the data collection. Eugene Milener programmed the stimuli and supervised the data collection for Experiment 3. Richard Burrignt was a valued source of intellectual stimulation. Reprint requests may be addressed to Robert A. Malmi, GTE Laboratories, Inc., 40 Sylvan Road, Waltham, MA 02254.

—Article accepted by previous editor, Robert A. Bjork

pairs of correlated numbers. In an attempt to make it difficult for subjects to employ such strategies as adding or subtracting the numerical stimuli to generate responses, the 100 pairs of stimuli were presented very rapidly (1 sec each). Furthermore, each number consisted of three digits. After 100 pairs were shown, 10 test stimuli which had not appeared before were shown. Each subject was asked to predict what the other member of the pair would be for each test stimulus. The contention is that the present experimental paradigm encourages the use of intuitive covariation judgments, rather than the creative but inappropriate number-manipulation strategies which were probably used by subjects in the previous research. The empirical question was whether the subjective predictions would reflect the correlation represented in the stimuli.

EXPERIMENT 1

The primary purpose of Experiment 1 was to map performance in this task as a function of the sign and degree of the covariation represented in the stimuli. In different experimental conditions, the correlations of the stimuli were .63, .41, .20, -.01, -.43, -.61, and -1.00.

Method

Design and Stimuli. The stimuli for the independent experimental conditions was a bivariate distribution of 100 three-digit numbers. Using standard conventions, the X variable will be considered to be the predictor of the Y variable. The X distribution was the same in all conditions and was generated by computer with the formula $X_i = M_x + S_x(E_i)$, where M_x is the mean of the X scores (400), S_x is the standard deviation (50), and E_i is a normal deviate having a value from zero to unity. The characteristics of the X distribution are given in Table 1. The Y distributions were generated to have a given correlation with the X distribution. This was done with the formula $Y_i = B(X_i - M_x) + M_y + S_y \sqrt{1 - B^2} (E_j)$, where M_y is the mean of the Y scores (400), S_y is the standard deviation of the Y scores (50), and B is the correlation of the X and Y scores. The characteristics of the seven distributions are also shown in Table 1.

The 10 test stimuli that were given to each subject for predicting the Y stimulus sampled the range of the numbers subjects saw, but they did not occur in the X distribution. These test stimuli were the same for all subjects and were 280, 326, 401, 456, 505, 294, 333, 396, 462, and 492. The order of the stimuli within halves (first five, second five) was randomized for each subject to facilitate assessment of testing effects, but since none were found in any of the experiments reported in this paper, this issue will not be mentioned further.

Procedure. Each subject sat alone in a cubicle about 50 cm away from a monitor on which the stimuli appeared. An Apple II+ generated the stimuli. Subjects entered their responses on a terminal. All subjects were asked to imagine that they were guidance counselors examining many pairs of aptitude test scores. The scores were called

the "left" and "right" scores. First a right score appeared; a short time later the left score appeared, then the screen was blanked and the next pair was shown. Subjects were asked to "get a general idea about the relationship between the two tests."

The left scores were the X distribution and the right scores were the Y distribution. The X score appeared on the screen first and after 500 msec the Y score appeared to the right of the X score. After 500 msec, the screen was blanked for 200 msec, after which another pair of X and Y scores was shown, and so on for 100 pairs. Each subject received a different random sequence of the pairs. Immediately after the last pair was shown, the test stimuli, which the subject was told were left scores, were shown one at a time. The subject was asked to "try to predict what the right score would be, on the average." The response to each test stimulus was typed on the terminal. The test was subject paced. Each subject received a randomized order of the test stimuli.

Subjects. One hundred forty-two students from general psychology classes at the State University of New York at Binghamton served in Experiment 1. Two students in Condition -1.00 noticed that if the XY stimuli for each pair were summed the result was a constant, and generated their responses accordingly. These 2 subjects were replaced. Subjects were assigned to conditions with a block-randomized schedule.

Results

Two major dependent measures were examined in Experiment 1, as well as in the subsequent experiments.

1. The best fitting line for the subjective estimates and the 10 test stimuli was calculated for each subject, using the least squares criterion. The degree to which the slope value of this line deviates from zero is a measure of the degree to which the subjective estimates reflect covariation based on inspection of the stimuli.

2. A second analysis determined what correlation coefficient would be associated with the line determined by the mean estimates given to each test stimulus and the test stimuli. This was accomplished by first rearranging the terms in the raw-score regression equation to solve for r . All terms in the equation have values in the experimental conditions (i.e., the means and standard deviations of the X and Y stimuli) except the value of the predictor (X_i) and the predicted value of Y given X_i (Y'). Let X_i be a test stimulus and let Y' be the associated point on the fitted line based on the subjective estimates. The calculations result in a correlation coefficient based on the subjective predictions given the test stimuli and the statistics of the stimuli subjects were shown. This derived measure will be called a subjective correlation coefficient.

The slope measures from the best fitting lines for the subjective estimates and the test stimuli in the seven experimental conditions were subjected to an analysis of variance in which the null hypothesis was rejected

Table 1
Characteristics of the Stimuli in Experiments 1 and 2

	Condition						
	.63	.41	.20	-.01	-.43	-.61	-1.00
Distribution	X	Y	Y	Y	Y	Y	Y
Mean	401.20	394.05	392.35	391.14	390.32	390.30	391.05
SD	45.52	45.83	47.19	48.44	49.58	50.50	50.15

Note—The X stimuli were the same in all conditions.

[$F(6,133) = 9.03$, $MSE = 0.15$]. A Newman-Keuls analysis revealed the following pattern, where conditions not sharing an underscore differed reliably ($p < .05$):

.63 .41 .20 - .01 - .43 - .61 -1.00

Performance in Conditions .63 and .41 will be considered together; that in Conditions .20 and - .01 together; that in Conditions - .43 and - .61 together; and performance in Condition -1.00 will be considered last.

Conditions .63 and .41. The mean estimates reflect a greater degree of covariation than existed in the stimuli. The mean slope value was .79 ($SD = .17$) in Condition .63 and was .60 ($SD = .22$) in Condition .41. The actual slope values of the stimuli were .63 and .43 in Conditions .63 and .41, respectively. The actual slope values were less than the 95% confidence limit based on the subjective slope measure in both conditions. The slope values associated with the best fitting line based on mean estimates were .77 and .60 for Conditions .63 and .41, respectively. When the stimuli reflected a correlation of .63, the subjective correlation was .69. When the stimuli were correlated .41, the subjective correlation was .60.

Conditions .20 and - .01. The estimates again tended to reflect a greater degree of covariation than existed in the stimuli. The mean slope value was .51 ($SD = .31$) in Condition .20 and was .55 ($SD = .33$) in Condition - .01. The actual slope values were obviously less than the 95% confidence limit based on the subjective slope measures. The slope values based on the mean estimates were .53 and .57 for Conditions .20 and - .01, respectively. When the stimuli were correlated .20, the subjective correlation was .41. When the stimuli were correlated - .01, the subjective correlation was .51.

Conditions - .43 and - .61. The subjective estimates did not reflect a negative relationship that was represented in the stimuli; rather, the relationship indicated by the mean estimates was slightly positive. The mean slope measure was .23 ($SD = .37$) in Condition - .43 and was .18 ($SD = .42$) in Condition - .61. The actual slope values were not within the appropriate confidence interval based on the subjective slope measures, of course. The slope values based on the mean estimates were .24 and .18 in Conditions - .43 and - .61, respectively. When the stimuli were correlated - .43, the subjective correlation was .17. When the stimuli were correlated - .61, the subjective correlation was .19.

Condition -1.00. The estimates reflected virtually a null relation of the XY stimuli even when the actual relationship was perfectly negative in a statistical sense. The mean slope measure was - .11 ($SD = .42$). The slope value for the line based on the mean estimates was - .12. When the stimuli reflected a correlation of -1.00, the subjective correlation was - .08. Inspection of the individual slope scores suggested pronounced bimodality with about equal numbers of scores being positive or negative. Only four scores were highly negative ($> .70$).

Discussion

To summarize the major results of Experiment 1, it seems that when performance in the seven experimental conditions is considered overall, one can conclude that the subjects were sensitive to variations in the degree of covariation in the stimuli. The support for this statement is that performance was different when the stimuli had mild positive, virtually null, mild negative, and perfectly negative degrees of covariation. Performance was different among the conditions, but was not always consistent with Pearsonian predictions.

When the stimuli reflected a positive correlation (Conditions .63 and .41), the subjective estimates generally reflected a greater degree of covariation than was represented in the stimuli. Although the subjective estimates were not perfectly accurate, it should be noted that most subjects gave estimates which reflected a generally appropriate degree of positive covariation.

When the correlation of the stimuli was virtually null (Conditions .20 and - .01), the tendency of most subjects was to give estimates that reflected a slightly positive degree of covariation. There are several possible reasons for this. One is that subjects may have a response bias to give estimates resembling the test stimuli. Another is that subjects may have "injected" an "illusory correlation" where none in fact existed (e.g., Hamilton & Gifford, 1976).

Most subjects were quite unable or unwilling to give estimates reflecting a negative correlation, even when the correlation was perfectly negative.

EXPERIMENT 2

In Experiment 2, the hypothesis examined was that the subjective estimates would reflect cognitive processes analogous to the principles of statistical theory. The basic procedures and stimuli were the same as those used in Experiment 1, but subjects were given a second trial. The prediction was that the subjective estimates would regress toward the regression function of the stimuli, due to more exposure to the stimuli. The overestimation of the degree of covariation in Conditions .63 and .41 should decrease, as should the "illusory correlation" in Condition - .01 on the second trial. It was also of interest to see whether further exposure to negatively correlated stimuli would result in more appropriate responses.

Method

Design and Procedure. Conditions .63, .41, - .01, - .43, and - .61 of Experiment 1 were included in Experiment 2. The procedures and stimuli were the same as those described for Experiment 1, except that after the estimates were made to the test stimuli, a message appeared on the monitor informing the subject that the same stimuli would be shown again and the same test given again. No feedback was given after the Trial 1 estimates. The stimuli were presented in a different random order on the second trial.

Subjects. One hundred students from the same source as before contributed data. They were assigned to conditions with a block-randomized schedule.

Results

Conditions .63 and .41. An analysis of variance using individual slope measures showed that performance changed very little on Trial 2 in Condition .63, but the slope of the best fitting line decreased on Trial 2 in Condition .43. The mean slope measures were .73 ($SD = .18$) and .78 ($SD = .19$) on Trials 1 and 2, respectively, in Condition .61, and were .70 ($SD = .18$) and .57 ($SD = .31$) on Trials 1 and 2, respectively, in Condition .43. The interaction of condition and trial was significant [$F(1,38) = 4.95$, $MSe = .032$]. The source of the interaction is the lack of any decrease in the slope measure in Condition .61 and the decrease in the slope measure in Condition .43 from Trial 1 to Trial 2. No other effects were significant, although the slope measures were generally higher in Condition .61 than in Condition .43 [$F(1,38) = 3.53$, $MSe = .068$, $p < .10$]. The actual slope measures did not fall within the 95% confidence limits in any case. The subjective correlation coefficients were .76 and .83 on Trials 1 and 2, respectively, in Condition .61, and were .73 and .57 on Trials 1 and 2, respectively, in Condition .43.

The distributions of slope measures looked similar (bell-shaped around the means reported above) on Trial 1 for both conditions. There was little change on Trial 2 in Condition .61, but in Condition .43 the distribution shifted to lower values on Trial 2. Eight of the 20 subjects in Condition .61 gave estimates reflecting lower slope measures on Trial 2 than on Trial 1. This was the case for 14 of the 20 subjects in Condition .43. In other words, 40% and 70% of the subjects in Conditions .61 and .43, respectively, gave estimates on Trial 2 that were in the appropriate direction, namely, regressing toward the regression function of the stimuli.

Conditions -.01, -.43, and -.61. These conditions did not differ reliably, there was no effect of trials, and there were no interactions for the individual slope measures. The mean slope measures were .27 ($SD = .34$) and .28 ($SD = .43$) on Trials 1 and 2, respectively, in Condition -.01; .30 ($SD = .38$) and .17 ($SD = .37$) in Condition -.43; and .14 ($SD = .47$) and .05 ($SD = .48$) in Condition -.61. The subjective correlations on Trials 1 and 2 were .27 and .26 in Condition -.01, .29 and .16 in Condition -.43, and .14 and .06 in Condition -.61.

The shape of the distribution of slope measures was bimodal in Condition -.01, and little change was evident between the two trials. Most slope values were low and positive on Trial 1 in Condition -.43, and there was little change on Trial 2. A shift in the distribution over trials was evident in Condition -.61, however. On Trial 1, 4 subjects gave estimates reflecting negative covariation; 8 subjects did so on Trial 2. Three of the 20 subjects had positive slope values on Trial 1 and negative slope values

on Trial 2. One subject had a slope value of $-.86$ on Trial 1, and a slope value of $.38$ on Trial 2.

Discussion

It appears that when confronted with positively correlated numerical stimuli, most subjects made estimates that were generally consistent with the principles of statistical theory. This was not the case when the stimuli had a null correlation. Furthermore, performance in Experiment 2 indicates that the subjects were utterly incapable of giving estimates that would reflect only the sign of negatively correlated stimuli. These points will be discussed in turn.

The estimates most subjects gave after viewing positively correlated numbers reflected positive covariation of a degree less than unity, so in that sense the estimates were regressive in nature. That is, most estimates in Conditions .43 and .63 in Experiments 1 and 2 were intermediate in value with respect to the test stimulus and the stimulus mean. This is always the case for statistical predictions unless the correlation is unity. In this general sense, then, the estimates were consistent with statistical theory. A most critical finding was that the estimates given in Condition .43 reflected regression toward the actual regression function on the second trial in Experiment 2, but this was not true of estimates in Condition .61. The regression of the estimates observed is consistent with the notion that the subjects had some intuitive knowledge of statistical theory. Several points will be made about this finding.

First, it is possible that the present experimental procedures did not have sufficient power to detect regression of the estimates given in Condition .61 in Experiment 2. Due to the greater initial overestimation of the covariation in Condition .43, relative to Condition .61, there was more "room" for regression to be observed in the former than in the latter condition.

Second, a similar argument can be stated in psychological terms. In a statistical sense, more data are required to detect low degrees of covariation than to detect higher degrees of covariation. For example, about 19 degrees of freedom are needed to decide that a correlation of .43 is not truly null, whereas 9 degrees of freedom are needed to make this decision for a correlation of .61. Perhaps an analogous principle applies to the cognitive processes underlying the estimates observed in Condition .43 in Experiment 2. The cognitive system may require progressively more data to know about progressively lower degrees of covariation.

The estimates given when the stimuli had virtually no degree of covariation (Condition .20, Experiment 1; Condition -.01, Experiments 1 and 2) tended to reflect slightly positive degrees of covariation. This may have been the result of demand characteristics. The instructions given to the subjects requested that they "get a general idea about the relationship of the test scores," so many subjects may have felt obligated to give estimates reflect-

ing a relationship even if they "knew" that none really existed. On the other hand, the tendency to endow a relationship between stimuli when none exists may be pervasive.

Most of the subjects exposed to negatively correlated numerical stimuli did not give estimates reflecting the sign, much less the degree, of covariation. The results of the experiments described below suggest that this inability is associated with numerical stimuli, and so is probably the result of some response bias unique to this type of stimuli.

EXPERIMENT 3

The major purpose of this experiment was to examine intuitive covariation estimation with nonnumerical stimuli. The stimuli in Experiment 3 were pairs of lines differing in length. After the presentation of 50 pairs of lines reflecting a given correlation of their lengths, the subjects were shown a series of lines and were asked to generate an appropriate other member of the pair based on the covariation represented in the stimuli. There were two major reasons for this procedure.

First, subjects' inability to give appropriate estimates after inspection of negatively correlated numerical stimuli may be the result of a response bias or other processes associated with numerical stimuli. Numbers may be unique stimuli for any of several reasons: a number is obviously a symbol, it has little meaning unless it is associated with something, and, at least for some people, numbers inspire negative affect. More appropriate estimates of covariation may be observed with nonnumerical stimuli.

Second, the major goal of this research was to study the kind of intuitive judgments about covariation which are presumably made routinely in everyday life. One might question the ecological validity of studies about covariation using numerical stimuli. Line lengths are perceptually salient and may better represent stimulus dimensions people typically encounter than do numerical stimuli.

Method

Design and Stimuli. Five between-subjects conditions were defined by the correlation represented by the variable lengths of 50 pairs of lines shown to subjects. The conditions are labelled in terms of the correlation of the stimuli: .59, .41, .01, -.41, and -.59. After the presentation of the 50 line pairs, 8 lines were presented, and for each of these the subject was asked to generate a line that would go with it, based on the stimuli just seen. Each subject was

then shown the 50 line pairs again, in a different order, and was tested again.

The line pairs were displayed with the low-resolution graphics of an Apple II+. The bivariate distributions of vertical lines had a mean of about 19 in these units and standard deviations of about 9 units. On the monitor used, the lines ranged from 0.30 cm to 10.70 cm. The test stimuli ranged from .85 cm to 10.15 cm.

Procedure. The subject was instructed to "try to get a general idea about the relationship between the left line's length and the right line's length." The left line appeared for 250 msec, then the right line appeared for 750 msec, then the screen was blanked for 400 msec; then the next left line appeared, and so on for 50 pairs of lines. After the 50 pairs of lines were shown, the subject was told that several "left lines" would be shown, and for each the subject was to generate a right line that reflected the relationship present in the stimuli. The response was made with the game paddle of the Apple II+. A second trial was given, in which the order of the stimuli was randomized again.

Subjects. One hundred students from the same source as before served. They were assigned to condition randomly.

Results

The mean subjective correlation coefficients for the experimental conditions are summarized in Table 2. In each condition, the subjective correlation coefficients seem to reflect adequately the correlation represented by the stimuli. Also, the subjective correlation coefficients seem to approach the objective correlation of the stimuli from Trial 1 to Trial 2 in each condition except Condition .01.

Conditions differed significantly [$F(4,95) = 36.25$, $MSe = .17$]. A Newman-Keuls analysis revealed the following pattern, where conditions sharing an underscore do not differ and otherwise do differ significantly:

.59 .41 .01 -.41 -.59

No other effects were significant. In a separate (ad hoc) analysis in which only Conditions .59 and .41 were included, a marginally significant ($p < .10$) interaction of condition and trial was found. This reflects the trend of the subjective correlation coefficient to increase in Condition .59 and to decrease in Condition .41 from Trial 1 to Trial 2. This trend reflects changes in the measure that more adequately reflect the correlation of the stimuli.

Discussion

The results of Experiment 3 show that the estimates reflected accurately the degree and sign of the correlation of a bivariate distribution of lines varying in length. Although the data do not allow the conclusion that subjects were able to discriminate between stimuli with correlations of .59 and .41 or -.59 and -.41, the trends are

Table 2
Subjective Correlation Coefficients for Experiment 3

	Condition									
	.59		.41		.01		-.41		-.59	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Trial 1	.42	.07	.45	.06	-.02	.09	-.25	.09	-.21	.09
Trial 2	.58	.06	.40	.08	.15	.08	-.31	.10	-.46	.08

suggestive that such discrimination may have been present. In each of these conditions, the estimates improved from Trial 1 to Trial 2 (although not statistically significantly) in that they reflected better the actual correlation represented by the stimuli.

The estimates subjects made after being shown negatively correlated stimuli accurately reflected the negative relationship by the stimuli. There appeared to be no difference in performance in conditions in which the sign of the correlation was positive or negative other than the sign of the relationship reflected by the estimates. In Experiments 1 and 2, estimates given in conditions in which the numerical stimuli were negatively correlated did not reflect the sign of the correlation. There are many differences between numbers and lines, of course, but it seems most plausible that a strong response bias associated with numerical stimuli may have affected performance in the conditions representing negative correlations in Experiments 1 and 2.

EXPERIMENTS 4 AND 5

A concern about ecological validity was noted earlier. One could argue that numerical and line stimuli do not represent adequately the kinds of stimulus dimensions important in the assessment of covariation which is thought to be important in everyday social interaction. The purpose of Experiments 4 and 5 was to examine the ability to assess covariation of stimuli varying on affective dimensions which presumably figure importantly in social cognition.

Which stimulus dimensions are most important, or at least prevalent, when an assessment of covariation may affect social cognition and behavior? Two dimensions that may be involved are pleasantness and familiarity. For example, one of the important judgments one makes upon meeting someone is whether the person is to be liked or not. Covariations between one's judgment of a person's pleasantness and one's judgment of his/her other characteristics affect social cognition. Similarly, familiarity of stimuli plays a vital role in clinical and medical judgments and in judgments of validity (Hasher, Goldstein, & Topino, 1977).

The stimulus dimensions of pleasantness and familiarity were manipulated in Experiments 4 and 5, respectively, by presenting a series of word-line pairs. Subjects were asked to think of the words in terms of their pleasantness (Experiment 4) or their familiarity (Experiment 5) and to determine the relationship of the word's attribute to the associated line, which varied in length. Numerical normative ratings for pleasantness and familiarity were used to generate line lengths associated with the words to represent a given correlation. After inspection of 100 word-line pairs, 10 new words were shown; for each, subjects were to generate a line length to represent the correlation of the stimuli.

METHOD

Design and Stimuli. In both Experiments 4 and 5, five between-subjects conditions were defined by the correlation of the words' ratings of pleasantness (Experiment 4) or familiarity (Experiment 5) and line length. The conditions are named in terms of the correlation: .60, .40, .00, -.40, and -.60. Two trials were given. Numerical normative ratings of pleasantness and familiarity were obtained for two sets of 100 words from Toggia and Battig (1978). Two sets of 10 new words were chosen as test stimuli. No word was included in more than one set. Using the same formula as given in Experiment 1, lines varying in length were generated for each word in the two sets so that a given correlation was represented between the word's rating and the line's length.

Procedure. Subjects were given practice thinking about the relevant dimensions for words. They were asked to rate 10 words for pleasantness (Experiment 4) or familiarity (Experiment 5). They were then asked to "try to determine the general relationship between the word's pleasantness (or familiarity) and the line's length." The 100 word-line pairs were displayed on the monitor at a rate determined by the subject (by pressing a terminal key), because pilot work showed that subjects responded at very different rates depending on the words' rating values (faster for words rated very high or low, slower for words having intermediate ratings). After the 100 word-line pairs were shown, the 10 test words appeared, one at a time; for each, subjects were to construct a line length, using the paddle. After the subjects had responded to the 10 words, the word-line pairs were randomized and a second trial was given.

Subjects. One hundred students from the same source as before served in Experiment 4, and another 90 served in Experiment 5. They were assigned to condition randomly.

Results

Experiment 4. The subjective correlation coefficients for each subject in the five conditions were subjected to an analysis of variance. Conditions differed reliably [$F(4,95) = 122.59$, $MSe = .12$]; no other effects were significant. A Newman-Keuls analysis revealed that Conditions .60 and .40 differed from Condition .00, which differed from Conditions -.40 and -.60, which did not differ. The mean subjective correlation coefficients for Conditions .60, .40, .00, -.40, and -.60 were .73, .67, .29, -.49, and -.58, respectively. The subjective correlation coefficients seem to match the correlation represented in the stimuli well when the overall pattern is examined.

Experiment 5. The same analysis was performed on the subjective correlation coefficients and the same pattern was obtained. Conditions differed reliably [$F(4,85) = 35.84$, $MSe = .32$], and no other factor was significant. Conditions .60 and .40 did not differ, but they differed from Condition .00, which differed from Conditions -.40 and -.60, which did not differ. The mean subjective correlation coefficients for Conditions .60, .40, .00, -.40, and -.60 were .74, .61, .30, -.30, and -.56, respectively.

Discussion

The purpose of Experiments 4 and 5 was to examine people's ability to assess the degree of covariation of an

affective stimulus dimension. Normative ratings for the pleasantness and familiarity of words were used to generate line lengths to represent a given correlation between the ratings and line lengths. Subjects generated lines to test words to reflect the correlation of the stimuli. For the most part, the responses accurately reflected the correlation represented in the stimuli. This was the case for both affective stimulus dimensions. These dimensions were examined because they presumably figure prominently in social cognition.

EXPERIMENT 6

Experiment 6 was an empirical extension of Experiments 4 and 5. In Experiment 6, subjects were asked to assess the covariation of two bivariate distributions simultaneously. Subjects were shown 100 sets of two word-line pairs, one pair representing covariation between the word's pleasantness and line length and the other pair representing covariation between the word's familiarity rating and line length.

There are many combinations of correlations possible when two bivariate distributions are involved; four possibilities are examined here.

1. Condition .60/.60. The correlation between the ratings and line lengths is set at .60 for both word-line sets. One possibility is that the subjective assessment of the degree of the two positive relationships may be mutually enhanced.

2. Condition .60/.00. The correlation between the pleasantness rating and line length is .60, and that between familiarity rating and line length is zero. If a positive relationship is detected for one pair, will it be "transferred" to the pair for which a null correlation exists?

3. Condition .60/-.60. The pairs based on pleasantness ratings are positively correlated and those based on familiarity ratings are negatively correlated. Will the difference in the sign of the correlation of the pairs enhance, interfere with, or have no effect on the assessment of covariation?

4. Condition .00/.00. The correlation is set at zero for both sets of word-line pairs.

Method

Design and Stimuli. Four between-subjects conditions were defined in terms of the stimuli shown to subjects. The stimuli used in Conditions .60, .60, .60, and .00 in Experiment 4 were associated with the stimuli used in Conditions .60, .00, -.60, and .00 in Experiment 5 to form the stimuli for Conditions .60/.60, .60/.00, .60/-.60, and .00/.00 in Experiment 6. The test words from Experiments 4 and 5 were also used for the appropriate stimuli in Experiment 6.

Procedure. Subjects were given practice in rating words for pleasantness and familiarity. They were asked to think of the words shown on the left on the monitor in terms of pleasantness and the words on the right on the monitor in terms of familiarity. They were asked to assess the general relationship between the words' attributes on these dimensions and the associated line lengths. The two word-line pairs were shown on the same line of the monitor, side by side. The subject pressed the space bar to see a new set of pairs. After 100 sets of pairs had been seen, the 10 test stimuli for the

pleasantness set were shown, one at a time, and a line length was generated for each. Then the 10 test words for the familiarity set were shown, and a line length was generated for each.

Subjects. Seventy-two students from the same source as before participated. They were assigned to condition randomly.

Results

It is useful to compare the subjective correlation coefficients observed in Experiment 6 with those observed in Experiments 4 and 5. In Condition .60 in Experiment 4, the mean subjective correlation was .73. The means for responses to the same stimuli in Experiment 6 were .68 ($SEM = .04$), .73 ($SEM = .05$), and .58 ($SEM = .07$). In Condition .60 in Experiment 5, the mean coefficient was .74. In Condition .60/.60, the mean for responses to the same stimuli was .64 ($SEM = .04$). In Condition -.60 in Experiment 5, the mean coefficient was -.56. The mean was -.42 ($SEM = .11$) for responses to the same stimuli in Condition .60/-.60. The means were .29 and .30 in Condition .00 in Experiments 4 and 5, respectively. The comparable means in Condition .00/.00 were .34 ($SEM = .09$) and .15 ($SEM = .12$), respectively.

The overall pattern of results leads to a conclusion that performance is not greatly different whether these covariation estimation tasks are presented singly or simultaneously.

GENERAL DISCUSSION

This research was prompted by a large number of studies whose results led some theorists (e.g., Nisbett & Ross, 1980) to conclude that people are utterly unable to assess covariation adequately. The importance of this conclusion is that the ability to assess covariation has been and is a critical element of theoretical approaches for the understanding of social behavior (Kelly, 1972, 1973). According to attribution theory, knowledge of the covariation between stimuli is a critical determinant of social behavior. The main message of the present research is that the conclusion stated by Nisbett and Ross (1980, p. 286), that "under most circumstances [the ability to assess covariation] requires mental operations that are simply beyond our intuitive, informal capacities," is not consistent with the results of the experiments reported herein. The present research shows that people can assess covariation adequately.

In Experiments 1 and 2, the stimuli were pairs of numbers. Three findings emerged. First, the estimates given in conditions in which the stimuli were negatively correlated did not reflect even the sign of the relationship. Second, the estimates reflected a mildly positive relationship when in fact none existed. Third, the estimates seemed to reflect the covariation of positively related stimuli quite well. In Experiment 2 there was evidence that the estimates improved from Trial 1 to Trial 2, even with no feedback.

In Experiment 3, the stimuli were pairs of lines which varied in length. The subjective estimates reflect relation-

ships quite accurately when the pattern of results is considered. This suggests that the difficulties people had with negatively correlated numerical stimuli are probably associated with some aspect of the stimuli.

The stimuli presented to subjects in Experiments 4, 5, and 6 were meant to represent stimulus dimensions presumed to figure importantly in everyday assessment of covariation, namely, the dimensions of pleasantness and familiarity. Once again, when the pattern of results in these experiments is considered, the assessment of covariation seems quite adequate and in some cases remarkably good.

Other studies (some mentioned in the introduction) have resulted in an opposite conclusion. Although this is speculation, it seems the methods and procedures in those studies may have allowed or encouraged inappropriate strategic processes. Slovic (1974), for example, reported that subjects in his study claimed to be adding, subtracting, and multiplying the numerical stimuli to produce a response. All subjects in the six experiments reported herein claimed that their responses were "guesses" and could not tell us how else they might have produced a response. These subjects were probably not able to employ inappropriate strategies due to the rate of stimulus presentation for the numerical stimuli in Experiments 1 and 2 and the nature of the stimuli in Experiments 3, 4, 5, and 6. Hence, subjects may be able to assess covariation well intuitively, but may falter if an inappropriate strategy interferes with intuitive processes.

The present research does not address the processes by which the estimates of covariation are made. Our previous research on how people estimate the average value of a category (Malmi & Samson, 1983) suggests that people have a representation of the distribution of stimuli in memory. It may be that people can represent bivariate distribution of stimuli and "fit" a linear function to relate the distributions of stimuli.

It is also interesting to speculate about the possibility that the ability to assess covariation could be improved by practice with feedback, using some of the tasks developed for this research. Perhaps such training would complement clinical training, in which the ability to assess covariation is critical.

In conclusion, statistical theory was developed to describe the phenomena of nature. Statistical tools are useful because they lead to generally appropriate decisions

and predictions. It would be surprising if people who have evolved in a world where statistical principles are useful did not have some intuitive knowledge of them. It seems that, indeed, they do have such intuitive knowledge.

REFERENCES

- BEYTH-MAROM, R. (1982). Perception of correlation reexamined. *Memory & Cognition*, **10**, 511-519.
- BIRNBAUM, M. H. (1976). Intuitive numerical prediction. *American Journal of Psychology*, **89**, 417-429.
- BREHMER, B. (1973a). Note on the relation between single-cue probability learning and multiple-cue probability learning. *Organizational Behavior & Human Performance*, **9**, 246-252.
- BREHMER, B. (1973b). Single-cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion. *Organizational Behavior & Human Performance*, **9**, 377-395.
- BREHMER, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior & Human Performance*, **11**, 1-27.
- HAMILTON, D. K., & GIFFORD, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis for stereotypic judgments. *Journal of Experimental Social Psychology*, **12**, 392-407.
- HASHER, L., GOLDSTEIN, D., & TOPPINO, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, **16**, 107-112.
- KAHNEMAN, D., & TVERSKY, A. (1982). On the study of statistical intuitions. In A. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.
- KELLEY, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner, (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.
- KELLEY, H. H. (1973). The process of causal attribution. *American Psychologist*, **28**, 107-128.
- MALMI, R. A., & SAMSON, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning & Verbal Behavior*, **22**, 547-559.
- NISBETT, R., & ROSS, L. (1980). *Human Inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- SHAKLEE, H., & MIMS, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Human Learning & Memory*, **8**, 208-224.
- SLOVIC, P. (1974). Hypothesis testing in the learning of positive and negative linear functions. *Organizational Behavior & Human Performance*, **11**, 368-376.
- SMELDELUND, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, **4**, 165-173.
- TOGLIA, M. P., & BATTIG, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

(Manuscript received December 31, 1984;
revision accepted for publication February 20, 1986.)